

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the total score distribution, and is derived by analysis of the total score distribution.

SUMMARY

result	No.	Score	Query Match	Length	DB ID	Description	%
c	1	45	4.4	3548	24	AAD38882 Alfalfa plastocyanin	
c	2	42.4	4.1	6078	22	AAT311861 Human KARP-1 DNA	
c	3	41	4.0	2614	19	AAV46621 Cotton promoter re	
c	4	40.6	3.9	1038	21	AAC627956 Senescence-like pr	
c	5	40.6	3.9	1038	24	ABK17061 Pinus radiata prom	
c	6	40.4	3.9	598	24	AAS15016 Rice anther-specific	
c	7	40.4	3.9	2096	21	O-methyl transfera	
c	8	40.4	3.9	2096	24	Eucalyptus grandis	
c	9	40.4	3.9	2611	24	Arabidopsis BONSAI	

PT of cDNAs -
 XX Example 2; Page 68-69; 74pp; English.
 CC The invention relates to a method for isolating and characterising an
 CC expression regulatory sequence for the expression of recombinant
 CC polypeptides and/or RNA. The method comprising producing at least one
 CC oligonucleotide primer from a cDNA library, where the
 CC upstream of the cDNAs. The method is useful for isolating,
 CC characterising and identifying a large number of known and unknown
 CC promoters that are active under any desired environmental condition to
 CC which a cell may be exposed and not just to the exemplified isolation of
 CC promoters that are capable of expression in specific conditions. The
 CC methods are also useful for cloning genes from any host, or from a
 CC specific tissue with such host, from which a cDNA library may be
 CC constructed; for the identification and isolation of analogous promoters,
 CC signal peptides and structural genes in several species of multicellular
 CC and unicellular organisms and as a high throughput identification system
 CC of candidate therapeutic targets. The promoter sequences may be used to
 CC regulate the synthesis of polypeptides. The present sequence is alfalfa
 CC plastocyanin gene coding fragment.

SQ Sequence 3548 BP; 1112 A; 576 C; 727 G; 1133 T; 0 other;
 XX Query Match 4 4%; Score 45; DB 24; Length 3548;
 Best Local Similarity 90.6%; Pred. No. 0.0029;
 Matches 48; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 AC AAF31861 standard; DNA; 6078 BP.
 AC AAF31861;
 XX DT 12-APR-2001 (first entry)
 DE Human KARP-1 DNA.
 XX Human; KARP-1; Ku86 autoantigen related protein; cancer; immune deficiency disorder; biliary tract cancer; leucine zipper protein; cytostatic; immunosuppressant; gene therapy; KARP-1 inhibitor; ds; Homo sapiens.
 XX US6171857-B1.
 PI Hendrickson EA;
 XX PD 09-JAN-2001.
 XX DR WPI; 2001-146208/15.
 XX DR P-PDB; AAB6590.
 PT Novel nucleic acids encoding leucine zipper protein, KARP-1 polypeptide, useful for treating cancer and immune deficiency disorder
 PT -
 XX Claim 1; Column 51-56; 61pp; English.
 CC The present sequence is given in a specification relating to an isolated
 CC Ku86 Autoantigen Related Protein (KARP-1) nucleic acid molecule. The

CC KARP-1 nucleic acid and KARP-1 protein are useful for the treatment
 CC and/or diagnosis of diseases such as cancer and immune deficiency
 CC disorders. They are useful in combination with a KARP-1 inhibitor that
 CC inhibits double stranded DNA base repair. Inhibitors of KARP-1 are
 CC useful in the diagnosis or treatment of conditions characterised
 CC by the loss of KARP-1 activity and in the treatment of cancer,
 e.g. biliary tract cancer.

SQ Sequence 6078 BP; 1806 A; 1301 C; 1392 G; 1543 T; 36 other;
 XX Query Match 4 4.1%; Score 42.4; DB 22; Length 6078;
 Best Local Similarity 97.7%; Pred. No. 0.028;
 Matches 43; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTTACTATAGGCAACGGTGGTCAGGCCGGCTGGATGAA 44
 Db 9 CTTACTATAGGCAACCGTGGTCAGGCCGGCTGGTCTGAA 52

RESULT 3
 AAV40621
 ID AAV40621 standard; DNA; 2614 BP.
 XX AAV40621;
 AC AAV40621;
 XX DT 26-OCT-1998 (first entry)
 DE Cotton promoter region from an expansin gene.
 XX KW Cotton fibre expansin promoter; transformation; transgenic; ss.
 XX OS Gossypium hirsutum
 XX PN WO9830698-A1.
 XX PD 16-JUL-1998.
 XX PF 07-JAN-1998; 98WO-US00151.
 XX PR 07-JAN-1997; 97US-0034914.
 XX PA (CALJ) CALGENE INC.
 PI Pear JR, Stalker DM;
 XX DR 1998-399143/34.
 XX PT New DNA containing the promoter of the cotton expansin gene - used
 PT for tissue-selective expression of genes that alter cotton fibre
 PT phenotype
 XX PS Claim 2; Fig 1a-d; 26pp; English.
 XX The present sequence represents an expansin promoter region from
 CC Gossypium hirsutum. This has been sequenced and found to be
 CC approximately 2200 bp in length, and is immediately 5' to the cotton
 CC fibre expansin coding region. The promoter can be used in cotton
 CC transformation by linking it to a gene of interest, thus enabling the
 CC modification of transgenic cotton fibre cells.
 XX SQ Sequence 2614 BP; 892 A; 436 C; 401 G; 882 T; 3 other;
 XX Query Match 4 0%; Score 41; DB 19; Length 2614;
 Best Local Similarity 100.0%; Pred. No. 0.048;
 Matches 41; conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTTACTATAGGCAACGGTGGTCAGGCCGGCTGGAT 41
 Db 83 CTTACTATAGGCAACCGTGGTCAGGCCGGCTGGTCTGAA 123

RESULT 4
 AAC62796

ID AAC62796	standard; DNA; 1038 BP.	PD 27-DEC-2001.
XX AC AAC62796;		XX
XX PF 02-FEB-2001 (first entry)		XX 20-JUN-2001; 2001WO-NZ000115.
XX DT 05-OCT-2000		XX 20-JUN-2000; 2000US-059401.
XX PR 24-FEB-2000; 2000WO-NZ00018.		PR 28-NOV-2000; 2000US-0724624.
DE Senescence-like protein promoter coding sequence #1.		XX (GENE-) GENESIS RES & DEV CORP LTD.
XX PA (FILET-) FLETCHER CHALLENGE FORESTS LTD.		PA (FILET-) FLETCHER CHALLENGE FORESTS IND LTD.
XX Perera R, Rice S, Eagleton C, Lasham A;		XX
XX OS Pinus radiata.		PI
XX PN WO20005B174-A1.		XX
XX PR 25-MAR-1999; 990S-0276599.		WPI; 2002-114583/15.
XX PR 30-JUL-1999; 990S-0146591.		DR p-PSDB; AAU0754.
XX PA (GENE-) GENESIS RES & DEV CORP LTD.		XX Novel polynucleotide promoter sequences from Pine and Eucalyptus useful
XX PA (FILET-) FLETCHER CHALLENGE FORESTS LTD.		PT for modifying expression of endogenous and/or heterologous
XX PI Perera R, Rice SJ, Eagleton CK;		PT polynucleotides in transgenic plants -
XX DR 2000-647236/62.		XX
XX DR 00005B174-A1.		PS Claim 1; Page 73-74; 121pp; English.
XX DR 28136.		XX The invention relates to isolated promoter sequences from <i>Pinus radiata</i> CC and <i>Eucalyptus grandis</i> , comprising a sequence chosen from leaf-, root-, CC flower-, pollen-, bud-, meristem-specific promoters or temporally regulated CC promoters such as xylogenesis-specific promoters. The promoter CC polypeptides and their related polynucleotides are useful in the CC production of genetic constructs, used for modifying gene expression in a CC target organism, in particular a plant. The method is useful for CC modifying expression of a polynucleotide that comprises an intron CC sequence, through removal of the intron sequence. The method is useful CC for modifying growth and development of plants, and cellular responses to CC external stimulus, such as environmental factors and disease pathogens. CC The sequences are useful in genome and physical mapping, in positional CC cloning of genes, in various assays to determine biological activity, to CC raise antibodies to isolate corresponding interacting proteins and other CC compounds, and to quantitatively determine levels of interacting proteins CC or other compounds. Sequences ABK17016-ABK17125 represent <i>Pinus radiata</i> CC and <i>Eucalyptus grandis</i> polynucleotides and PCR primers used in the method CC of the invention.
XX DR 2000-647236/62.		XX
XX DR 00005B174-A1.		SQ Sequence 1038 BP; 298 A; 220 C; 233 G; 287 T; 0 other;
XX DR 28136.		Query Match 3.9%; Score 40.6; DB 21; Length 1038;
XX DR 2000-647236/62.		Best Local Similarity 71.1%; Pred. No. 0.037; Pred. No. 0.037;
XX DR 00005B174-A1.		Matches 69; Conservative 0; Mismatches 24; Indels 4; Gaps 1; Gaps 1;
XX DR 28136.		XX
XX DR 2000-647236/62.		QY 2 TTACTATAGGCCACGGCGTGTGTCAGGCCGGCTGGTATGAAGGTGGAAACCTCAGTGG 61
XX DR 00005B174-A1.		Db 4 TTACTATAGGCCACGGCGTGTGTCAGGCCGGCTGGTATGAAGGTGGAAACCTCAGTGG 59
XX DR 28136.		QY 62 ATGCATAATCTGCTAGAGATACATCAATTCACA 98
XX DR 2000-647236/62.		Db 60 CTGTGTAAATTCTGCTAGTCACCCATAACTTTCTCA 96
XX DR 28136.		RESULT 5
XX DR 2000-647236/62.		ABK17061 standard; cDNA; 1038 BP.
XX DR 28136.		XX ABK17061
XX DR 2000-647236/62.		XX ID AAS15016 standard; cDNA; 598 BP.
XX DR 28136.		XX AC AAS15016;
XX DR 2000-647236/62.		XX DT 14-FEB-2002 (first entry)
XX DR 28136.		XX DE Rice anther-specific promoter #2.
XX DR 2000-647236/62.		XX DE Rice; promoter; gene therapy; transcription; monocot; anther;
XX DR 28136.		XX KW chromosome mapping; gene mapping; antisense technology;
XX DR 2000-647236/62.		XX KW plant genetic engineering; ss.
XX DR 28136.		XX OS <i>Oryza sativa</i> .
XX DR 2000-647236/62.		XX PN WO200198485-A1.
XX DR 28136.		XX PN WO200198485-A1.

PD 01-NOV-2001.
 XX
 PF 25-APR-2001; 2001WO-US13544.
 XX
 PR 26-APR-2000; 2000US-199870P.
 PR 12-JUL-2000; 2000US-217893P.
 PR 13-AUG-2000; 2000US-218366P.
 PR 23-AUG-2000; 2000US-227231P.
 PR 03-OCT-2000; 2000US-337738P.
 PR 29-NOV-2000; 2000US-253928P.
 XX
 PA (AKKA-) AKRADIX CORP.
 XX
 PI Perera JR, Lu M, Ray A;
 XX
 DR WPI; 2002-041419/05.
 XX
 PT Rice promoter sequences (1) useful in plant genetic engineering and
 molecular biology studies
 XX
 PS Claim 1; Fig 1B; 47pp; English.
 XX
 CC The invention relates to plant polynucleotide sequences from rice which
 encode promoter components of the cellular activation and transcription
 apparatus (5' cis regulatory DNA sequences). The sequences may be used in
 the modification of gene activation and/or expression in eukaryotes,
 especially rice and other monocots. They have many applications in
 standard molecular biology, e.g., as for chromosome and gene mapping, in
 polymerase chain reaction (PCR) technology, for the study of gene
 function and expression in vivo, specific down-regulation of target
 genes, and in the production of sense and/or antisense nucleic acids.
 CC The present sequence represents rice anther-specific promoter #2
 CC as described in the invention.
 XX
 SQ Sequence 598 BP; 174 A; 131 C; 113 G; 179 T; 1 other;
 XX
 Query Match 3.9%; Score 40.4; DB 24; Length 598;
 Best Local Similarity 75.8%; Pred. No. 0.031;
 Matches 50; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
 Qy 4 ACTATAGGCACGGTGTGTCAGGCCGGCTGGTAGGGTGGAACTCACTGGAT 63
 Db 1 ACTATAGGCACGGTGTGTCAGGCCGGCTGGTAGGGTGGATCAAATGTCGAT 60
 Qy 64 GCATAT 69
 Db 61 TAATAT 66
 XX
 RESULT 7 AAC62810
 ID AAC62810 standard; DNN; 2096 BP.
 XX
 AC AAC62810;
 XX
 DT 02-FEB-2001 (first entry)
 XX
 DE O-methyl transferase promoter coding sequence #2.
 XX
 KW Promoter; eucalyptus; pine; gene transcription; ds.
 OS Eucalyptus grandis.
 XX
 PN WO200058474-A1.
 PD 05-OCT-2000.
 XX
 PF 24-FEB-2000; 2000WO-NZ00018.
 XX
 PR 25-MAR-1999; 99US-0276599.
 PR 30-JUL-1999; 99US-0146591.
 XX
 PA (GENE-) GENESIS RES & DEV CORP LTD.

PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
 XX
 PI Perera R, Rice SJ, Bagleton CK;
 XX
 DR WPI; 2000-641236/62.
 DR P-PSDB; AAB2142.
 XX
 PT Novel promoter sequences useful for modulating transcription of plant
 DNA sequences of interest and production of polypeptides -
 PT
 PS Claim 1; Pages 61-62; 93pp; English.
 XX
 CC The present invention relates to promoter sequences from eucalyptus
 CC and pine. The present sequence is one such promoter. This sequence is
 useful for modulating the transcription of DNA sequences of interest. The
 CC sequences may also be used to tag or identify an organism or its
 CC reproductive material.
 XX
 SQ Sequence 2096 BP; 537 A; 535 C; 539 G; 485 T; 0 other;
 XX
 Query Match 3.9%; Score 40.4; DB 21; Length 2096;
 Best Local Similarity 97.6%; Pred. No. 0.05;
 Matches 41; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Qy 2 TTACTATAGGCACGGTGTGTCAGGCCGGCTGGATGA 43
 Db 3 TTACTATAGGCACGGCNGGTGTCAGGCCGGCTGGTGTGA 44
 XX
 RESULT 8 ABK17075
 ID ABK17075 standard; cDNA; 2096 BP.
 XX
 AC ABK17075;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Eucalyptus grandis promoter polynucleotide #31.
 XX
 KW Promoter; pine; leaf; flower; pollen; bud; meristem; xylogenesis;
 KW temporally regulated promoter; Pinus radiata; Eucalyptus grandis; ss;
 KW PCR primer.
 XX
 OS Eucalyptus grandis.
 XX
 PN WO200198485-A1.
 XX
 PD 27-DEC-2001.
 XX
 PF 20-JUN-2001; 2001WO-NZ00115.
 XX
 PR 28-NOV-2000; 2000US-0598401.
 PR 28-NOV-2000; 2000US-0744624.
 XX
 PA (GENE-) GENESIS RES & DEV CORP LTD.
 PA (FLET-) FLETCHER CHALLENGE FORESTS IND LTD
 XX
 PI Perera R, Rice S, Bagleton C, Lasham A;
 XX
 DR WPI; 2002-114583/15.
 DR P-PSDB; AAU80760.
 XX
 PT Novel polynucleotide promoter sequences from Pine and Eucalyptus useful
 PT for modifying expression of endogenous and/or heterologous
 PT polynucleotides in transgenic plants -
 PT
 PS Claim 1; Page 78; 121pp; English.
 XX
 CC The invention relates to isolated promoter sequences from Pinus radiata
 CC and Eucalyptus grandis, comprising a sequence chosen from leaf, root-,
 CC flower-, pollen-, bud-, meristem specific promoters or temporally
 CC regulated promoters such as xylogenesis-specific promoters. The promoter
 CC polypeptides and their related polynucleotides are useful in the
 PA (GENE-) GENESIS RES & DEV CORP LTD.

production of genetic constructs, used for modifying gene expression in a target organism, in particular a plant. The method is useful for modifying expression of a polynucleotide that comprises an intron sequence, through removal of the intron sequence. The method is useful for modifying growth and development of plants, and cellular responses to external stimulus, such as environmental factors and disease pathogens. The sequences are useful in genome and physical mapping, in positional cloning of genes, in various assays to determine biological activity, to raise antibodies, to isolate corresponding interacting proteins and other compounds, and to quantitatively determine levels of interacting proteins or other compounds. Sequences ABK17016-ABK1725 represent Pinus radiata and Eucalyptus grandis polynucleotides and PCR primers used in the method of the invention.

Sequence 2096 BP: 537 A: 535 C: 539 G: 485 T: 0 other:

Query	Match	Similarity	Score	DB	Length	209
	Best Local Matches	97.6%	40.4;	DB 24;		
	Local Conservative Matches	97.6%	0.065;	Fred. No.		
		41	0	Mismatches	1;	Indels
Y	2	T T A C T A T A G G G C A C G G T G G T C G A C G C C G G G C G T G T A T G A	4.3			
	1					
Z	3	T T A C T A T A G G G C A C G G T G G T C G A C G C C G G G C G T G T A T G A	4.4			
	2					

WO20020097-A2 .
 X 03 -JAN-2002 .
 X 25 -JUN-2001; 2001WO-US20172 .
 F 23 -JUN -2000; 2000US 213863P .
 R (WHED) WHITEHEAD INST BIOME
 X Hua J , Grisafi P , Fink GR ;
 I WPI; 2002-139899/18 .
 R

The present sequence is that of the BONSAI (BON1) gene of Arabidopsis thaliana. The coding region of the BON1 gene is given in ABA91256. The BON1 gene permits wild-type Arabidopsis plants to maintain a relatively constant size over a wide range of temperatures. Thus, bon1 null mutants produce miniature fertile plants at 22 degrees C, but a wild-type phenotype at 28 degrees C. BON1 has a direct role in regulating cell expansion and cell division at temperatures lower than those at which Arabidopsis is normally grown. The BON1 protein (see AAM50564) contains a Ca²⁺-dependent phospholipid binding domain and is associated with the plasma membrane. BON1 belongs to the copine gene family, which is conserved from protozoa to humans. The invention is directed to

isolated BON1, BON2, BON3, BAP1 and BAL nucleic acids (see ABA19126-60), which encode proteins (see AAM50644-48) that are necessary for normal growth, controlling cell expansion and cell division, affecting the size and rate at which the plant grows when exposed to lower temperatures. Transgenic plants are provided that are smaller than the wild-type as a result of inhibition of BON1, BON2, BON3, BAP1 and/or BAL, especially angiosperms and gymnosperms ornamental plants and turfgrass. Transgenic plants are also provided that are larger than the wild-type as a result of enhancement of BON1, BON2, BON3, BAP1 and/or BAL, especially crop plants and biomass plants. Modulation of these genes provides increased yield, or growth at a higher altitude or lower temperature.

Sentence 3611 BP: 809 A: 429 C: 498 G: 866 T: 9 other:

Query	Match	Score	DB	Length
Best Local Matches	97.9% ; 41;	0.074 ;	24 ;	2611;
Conservative Matches	97.9% ; 41;	0 ;	24 ;	2611;
1	CTTACATAGGCACCGGTGGTGAGGCCGGGCTGTATG	4.2		
2603	CTTACATAGGCACGGCTGGTGAGGCCGGGCTGTATG	2562		

SULT 10
C62780 AAC62780 standard; DNA; 411 BP.
AAC62780;
02-FEB-2001 (first entry)
Flower specific promoter coding sequence #2.
Promoter; eucalyptus; pine; gene transcription; ds.
Eucalyptus grandis.

05-OCT-2000.
24-FEB-2000; 2000WO-NZ00018.
25-MAR-1999; 99US-0276599.
30-JUL-1999; 99US-0146591.
(GENE-) GENESIS RES & DEV CORP LTD.
(FLET-) FLETCHER CHALLENGE FORESTS LTD
Perera R, Rice SJ, Eagleton CK;

Novel promoter sequences useful for modulating transcription of plant DNA sequences of interest and production of polypeptides -
Claim 1: Page 53; 93pp; English.

The present invention relates to promoter sequences from eucalyptus and pine. The present sequence is one such promoter. This sequence is useful for modulating the transcription of DNA sequences of interest. Sequences may also be used to tag or identify an organism or its reproductive material.

Sequence 411 BP; 112 A; 118 C; 80 G; 101 T; 0 other;

Query Match	3.98;	Score 40.2;	DB 21;	Length 411;
Best Local Similarity	93.38;	Pred. No. 0.029;		
Matches	42;	Conservative	0;	Mismatches 3;
				Indels 0;
				Gaps

4 ACTATAGGCAGGGTGTGCAAGGCCGGCTGTTATGAAGCTG 48

Db 1 ACTATAGGGCACCGTGTGTCAGGCCGGCTGGCTGAACTG 45 AAX35652 standard; cDNA; 594 BP.
 ID AAX35652
 XX AAX35652
 AC AAX35652;
 RESULT 11
 ABK17045 standard; cDNA; 411 BP.
 XX DT 09-JUL-1999 (first entry)
 AC DE 5' region of human heparanase cDNA.
 XX HEparanase; hpa; modulator; heparin-binding growth factor;
 XX cellular response; cytokine; cell interaction; plasma lipoprotein;
 XX cellular susceptibility; infection; disintegration;
 DE KW atherosclerotic plaque; wound healing; angiogenesis; restenosis;
 KW atherosclerosis; inflammation; neurodegenerative disease; neutralise;
 KW plasma heparin; micrometastasis; autoimmune lesion; renal failure;
 KW ss.
 KW PCR primer.
 XX OS Homo sapiens.
 XX PN WO9911798-A1.
 XX WO9911798-A1.
 XX PD 11-MAR-1999.
 XX PR 31-AUG-1998; 98WO-US17954.
 XX PF 02-JUL-1998; 98US-0109386.
 XX PR 02-SEP-1997; 97US-0922170.
 XX (FRIE/) FRIEDMAN M M.
 PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
 PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.
 XX PI Feinstein E, Pecker I, Vlodavsky I;
 XX DR 1999-30225/5.
 XX PT New human polynucleotide useful for treating angiogenesis,
 PT restenosis, and inflammation
 XX PS Example 8; Page 69; 63pp; English.
 XX The specificity describes a polypeptide having heparanase (hpa)
 CC activity. The recombinant protein is used as a modulator of
 CC heparin-binding growth factors, cellular responses to heparin-binding
 CC growth factors and cytokines, cell interaction with plasma lipoproteins,
 CC cellular susceptibility to viral, protozoal and bacterial infections
 CC or disintegration of neurodegenerative plaques. Heparanase may be
 CC useful for conditions such as wound healing, angiogenesis, restenosis,
 CC atherosclerosis, inflammation, neurodegenerative diseases, and viral
 CC infections. Mammalian heparanase can be used to neutralize plasma
 CC heparin, and anti-heparanase antibodies may be applied for
 CC immunodetection and diagnosis of micrometastases, autoimmune lesions,
 CC and renal failure in biopsy specimens, plasma samples, and body fluids.
 CC The present sequence represents the 5' sequence of human heparanase
 CC cDNA.
 XX SQ Sequence 594 BP; 133 A; 134 C; 209 G; 118 T; 0 other;
 Query Match 3.9%; Score 40; DB 20; Length 594;
 Best Local Similarity 100.0%; Pred. No. 0.041;
 Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 2 TTACTATAGGGCACCGTGTGTCAGGCCGGCTGGCTGAT 41
 Qy 4 ACTATAGGGCACCGTGTGTCAGGCCGGCTGGCTGGCTGAT 48
 Best Local Similarity 93.3%; Pred. No. 0.049;
 Matches 42; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 Db 1 ACTATAGGGCACCGTGTGTCAGGCCGGCTGGCTGGCTGAACTG 45
 Qy 5 ACTATAGGGCACCGTGTGTCAGGCCGGCTGGCTGGCTGGCTGAT 41
 Best Local Similarity 93.3%; Pred. No. 0.049;
 Matches 42; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 Db 2 TTACTATAGGGCACCGTGTGTCAGGCCGGCTGGCTGGCTGGCTGAT 41
 RESULT 13
 AAA75054 standard; cDNA; 594 BP.
 ID AAA75054
 XX AC AAA75054;
 XX AC AAA75054;

CC driving transcription in a seed-prefferred manner and can be used in an
CC expression cassette, to stably transform plant cells. The expression
CC cassette can be used to modify the fatty acid content of seeds, alter
CC the starch or carbohydrate profile, and/or alter the amino acid content
CC of the seed. It can also be used to deliver genes encoding important
CC traits for agronomics, including insect resistance, disease resistance,
CC herbicide resistance, and grain characteristics.

XX Sequence 1224 BP; 378 A; 239 C; 182 G; 425 T; 0 other;

Query Match 3.9%; Score 39.8; DB 21; Length 1224;
Best Local Similarity 95.3%; Pred. No. 0.074;
Matches 41; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 3 TACATAGGGCACCGTGGTCAGGGCCGGCTGGTATGAAAG 45
Db 1 TACATAGGGCACGGTGGTCAGGCCGGCTGGTGTAAAAG 43

Search completed: December 27, 2002, 22:42:34
Job time : 256 secs

Result No.	Score	Query Match	Length	DB ID	Description
1	42.4	4.1	60718	4 US-09-173-914-1	Sequence 1, Appli
2	39.8	3.9	409424	283-6	Sequence 6, Appli
3	39.6	3.8	7218	1 US-08-1232-463-14	Sequence 14, Appli
4	39	3.8	4526	4 US-09-323-195A-1	Sequence 1, Appli
5	39	3.8	4526	4 US-09-424-283-7	Sequence 7, Appli
c	38.4	3.7	921	4 US-09-137-648-4	Sequence 4, Appli
7	38.4	3.7	3791	4 US-09-570-367C-1	Sequence 1, Appli
c	38	3.7	336	4 US-09-276-599-13	Sequence 13, Appli
c	38	3.7	763	4 US-09-76-599-14	Sequence 14, Appli
10	37.8	3.7	565	4 US-09-323-195A-5	Sequence 5, Appli
c	37.8	3.7	1924	4 US-09-424-283-5	Sequence 5, Appli
c	37.4	3.6	48	4 US-08-913-014A-18	Sequence 18, Appli
c	37.4	3.6	2791	4 US-09-570-367C-1	Sequence 1, Appli
c	37.2	3.6	1478	4 US-09-545-814-28	Sequence 28, Appli
c	37.2	3.6	1478	4 US-09-545-814-30	Sequence 30, Appli
c	37.2	3.6	1677	4 US-09-545-814-13	Sequence 13, Appli
c	37.2	3.6	1677	4 US-09-545-814-15	Sequence 15, Appli
c	37.2	3.6	1749	4 US-09-545-814-4	Sequence 4, Appli
c	37.2	3.6	1749	4 US-09-545-814-6	Sequence 6, Appli
c	37.2	3.6	1919	4 US-09-545-814-31	Sequence 31, Appli
c	37.2	3.6	1919	4 US-09-545-814-33	Sequence 33, Appli
c	37.2	3.6	2610	4 US-09-545-814-1	Sequence 1, Appli
c	37.2	3.6	2610	4 US-09-545-814-3	Sequence 3, Appli
c	37	3.6	340	4 US-09-323-195A-3	Sequence 3, Appli
c	36	3.5	255	4 US-09-257-583-5	Sequence 5, Appli
c	36	3.5	342	4 US-09-323-195A-6	Sequence 6, Appli
27	3.5	555	4 US-08-905-223-37	Sequence 37, Appli	

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model
Run on: December 28, 2002, 00:01:58 ; Search time 54 Seconds
(without alignments)
5849.575 Million cell updates/sec

Title: US-09-945-376-3
Perfect score: 1030
Sequence: 1 cttactataggcacgcgtg. tgrtatataccaaaggcttgg 1030

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Issued_Patents_NA:*

1: /cgn2_6/podata/1/ina/5A_COMB.seq:*

2: /cgn2_6/podata/1/ina/5B_COMB.seq:*

3: /cgn2_6/podata/1/ina/6A_COMB.seq:*

4: /cgn2_6/podata/1/ina/6B_COMB.seq:*

5: /cgn2_6/podata/1/ina/PCTUS_COMB.seq:*

6: /cgn2_6/podata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

8

ALIGNMENTS

RESULT 1
US-09-173-914-1
; Sequence 1, Application US/09173914
; Patent No. 6171857
; GENERAL INFORMATION:
; APPLICANT: Hendrickson, Eric
; TITLE OF INVENTION: A New Method of Regulating DNA Dependent Protein Kinase Activity
; FILE REFERENCE: B0877//017/HK
; CURRENT APPLICATION NUMBER: US/09/173, 914
; CURRENT FILING DATE: 1998-10-16
; EARLIER APPLICATION NUMBER: 60/064, 557
; EARLIER FILING DATE: 1997-10-17
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO: 1
; LENGTH: 6078
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (178)...(178)
; NAME/KEY: unsure
; LOCATION: (230)...(230)
; NAME/KEY: unsure
; LOCATION: (232)...(232)
; NAME/KEY: unsure
; LOCATION: (234)...(234)
; NAME/KEY: unsure
; LOCATION: (453)...(453)
; NAME/KEY: unsure
; LOCATION: (473)...(473)
; NAME/KEY: unsure
; LOCATION: (610)...(610)
; NAME/KEY: unsure
; LOCATION: (612)...(612)
; NAME/KEY: unsure
; LOCATION: (2175)...(2175)
; NAME/KEY: unsure
; LOCATION: (1014)...(1014)
; NAME/KEY: unsure
; LOCATION: (173-914-1

1 CTTACTATAGGCCACGGTGGTCGAAGCCGGGGGTGATGAA 44
9 CTTACTATAGGCCACGGTGGTCGAAGCCGGGGGTGATGAA 52

```

RESULT 2
US-09-444-283-6
; Sequence 6, Application US/09444283
; Patent No. 6437219
; GENERAL INFORMATION:
; APPLICANT: Grimes, et al.
; TITLE OF INVENTION: Sucrose binding proteins
; FILE REFERENCE: 4630-50206
; CURRENT APPLICATION NUMBER: US/09/424-283
; PRIORITY FILING DATE: 1999-11-19
; PRIOR FILING DATE: 1998-05-21
; PRIOR APPLICATION NUMBER: US 60/047,568
; PRIORITY FILING DATE: 1997-05-22
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 6
; LENGTH: 3718
; TYPE: DNA
; ORGANISM: Glycine max
US-09-424-283-6

Query Match 3.9%; Score 39.8; DB 4; Length 3718;
Best Local Similarity 95.3%; Pred. No. 0.021; Indels 0; Gaps 0;
Matches 41; Conservative 0; Mismatches 2; Gaps 0;

Qy 1 CCTACTATAGGGCACCGCTGTCGAGCCCGGCCGCTGGTATGA 43
Db 107 CCTACTATAGGGCACCGCTGTCGAGCCCGGCCGCTGGTATGA 149

RESULT 3
US-08-232-463-14
; Sequence 14, Application US/08232463
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEFFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22213-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE: 26-AUG-1991
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE: 26-AUG-1991
; APPLICATION NUMBER: EP 91 114 300.6
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-3300
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
```

; PRIOR APPLICATION NUMBER: PCT/US98/10465
 ; PRIOR FILING DATE: 1998-05-21
 ; PRIOR APPLICATION NUMBER: US 60/047,568
 ; PRIOR FILING DATE: 1997-05-22
 ; NUMBER OF SEQ ID NOS: 15
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO: 7
 ; LENGTH: 4526
 ; TYPE: DNA
 ; ORGANISM: Glycine max
 US-09-424-283-7

Query Match 3.88; Score 39; DB 4; Length 4526;
 Best Local Similarity 100.0%; Pred. No. 0.042;
 Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 TTRACTATGGCAGCGCTGGTGACGCCGGCGGTGA 40
 Db 97 TTRACTATGGCACGCGTGGTGACGCCGGCGGTGA 135

RESULT 6

US-09-377-648-4/c
 ; Sequence 4, Application US/09377648
 ; Patent No. 6225529
 ; GENERAL INFORMATION:
 ; APPLICANT: Lappégaard, Kathryn
 ; TITLE OF INVENTION: Seed-preferred Promoters
 ; FILE REFERENCE: 0869
 ; CURRENT APPLICATION NUMBER: US/09/377,648
 ; CURRENT FILING DATE: 1999-08-19
 ; EARLIER APPLICATION NUMBER: US 60/097,233
 ; EARLIER FILING DATE: 1998-08-20
 ; SOFTWARE: FastSEQ for Windows Version 3.0
 ; SEQ ID NO 4
 ; LENGTH: 921
 ; TYPE: DNA
 ; ORGANISM: Zea mays
 ; FEATURE:
 ; NAME/KEY: promoter
 ; LOCATION: (1)...(922)
 US-09-377-648-4

Query Match 3.78; Score 38.4; DB 4; Length 921;
 Best Local Similarity 97.5%; Pred. No. 0.028;
 Matches 39; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTTACTATGGCAGCGCTGGTGACGCCGGCGGTGA 40
 Db 912 CTCACTATGGCAGCGCTGGTGACGCCGGCGGTGA 873

RESULT 7

US-09-570-367C-1
 ; Sequence 1, Application US/09570367C
 ; Patent No. 6338831
 ; GENERAL INFORMATION:
 ; APPLICANT: Gorczyński, Reginald M.
 ; TITLE OF INVENTION: Methods and Compositions for Immunomodulation
 ; FILE REFERENCE: 9579-21
 ; CURRENT APPLICATION NUMBER: US/09/570,367C
 ; CURRENT FILING DATE: 2000-05-05
 ; PRIOR APPLICATION NUMBER: US 60/064,764
 ; NUMBER OF SEQ ID NOS: 22
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 1
 ; LENGTH: 2791
 ; TYPE: DNA
 ; ORGANISM: Mus musculus
 US-09-570-367C-1

Query Match 3.7%; Score 38.4; DB 4; Length 2791;
 Best Local Similarity 87.5%; Pred. No. 0.051;
 Matches 42; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 4 ACTATAGGCACCGGTGGTCGACGCCGGCGGTGAAGGTGGAA 51
 Db 1 ACTATAGGCACCGTGGTCGACGCCGGCGTGTACTGAGAGAA 48

RESULT 8

US-09-276-599-13/c
 ; Sequence 13, Application US/09276599
 ; Patent No. 6380459
 ; GENERAL INFORMATION:
 ; APPLICANT: Perera, J. Ranjan
 ; TITLE OF INVENTION: Composition and methods for the
 ; FILE REFERENCE: 11000.1036
 ; CURRENT APPLICATION NUMBER: US/09/276,599
 ; CURRENT FILING DATE: 1999-03-25
 ; NUMBER OF SEQ ID NOS: 21
 ; SOFTWARE: FastSEQ for Windows Version 3.0
 ; SEQ ID NO 13
 ; LENGTH: 336
 ; TYPE: DNA
 ; ORGANISM: Pinus radiata
 US-09-276-599-13

Query Match 3.7%; Score 38; DB 4; Length 336;
 Best Local Similarity 100.0%; Pred. No. 0.022;
 Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 TTACTATAGGCACCGTGGTCGACGCCGGCGGTGTGA 39
 Db 334 TTACTATAGGCACCGTGGTCGACGCCGGCGGTGTGA 297

RESULT 9

US-09-276-599-14/c
 ; Sequence 14, Application US/09276599
 ; Patent No. 6380459
 ; GENERAL INFORMATION:
 ; APPLICANT: Perera, J. Ranjan
 ; TITLE OF INVENTION: Composition and methods for the
 ; FILE REFERENCE: 11000.1036
 ; CURRENT APPLICATION NUMBER: US/09/276,599
 ; CURRENT FILING DATE: 1999-03-25
 ; NUMBER OF SEQ ID NOS: 21
 ; SOFTWARE: FastSEQ for Windows Version 3.0
 ; SEQ ID NO 14
 ; LENGTH: 763
 ; TYPE: DNA
 ; ORGANISM: Pinus radiata
 US-09-276-599-14

Query Match 3.7%; Score 38; DB 4; Length 763;
 Best Local Similarity 100.0%; Pred. No. 0.034;
 Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 TTACTATAGGCACCGTGGTCGACGCCGGCGGTGTGA 39
 Db 761 TTACTATAGGCACGGTGGTCGACGCCGGCGGTGTGA 724

RESULT 10

US-09-323-195A-5
 ; Sequence 5, Application US/09323195A
 ; Patent No. 6462257
 ; GENERAL INFORMATION:

;

APPLICANT: Pullman, Gerald
; APPLICANT: Cairney, John
; APPLICANT: Perrier, Raniar
; TITLE OF INVENTION: VICILIN-LIKE SEED STORAGE PROTEIN GENE PROMOTER AND METHODS OF USING THE SAME
; FILE REFERENCE: IPST0009
; CURRENT APPLICATION NUMBER: US/09/323,195A
; CURRENT FILING DATE: 1998-06-01
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 565
; TYPE: DNA
; ORGANISM: Pinus taeda
; FEATURE:
; OTHER INFORMATION: n at 489 is a, c, g, or t
; OTHER INFORMATION: n at 503 is a, c, g, or t
; OTHER INFORMATION: n at 504 is a, c, g, or t
; OTHER INFORMATION: n at 522 is a, c, g, or t
; OTHER INFORMATION: n at 533 is a, c, g, or t
; OTHER INFORMATION: n at 543 is a, c, g, or t
; OTHER INFORMATION: n at 549 is a, c, g, or t
; OTHER INFORMATION: n at 564 is a, c, g, or t
US-09-323-195A-5

Query Match 3.7% Score 37.8; DB 4; Length 565;
Best Local Similarity 95.1%; Pred. No. 0.033; 0; Gaps 0;

Qy 4 ACTATAGGCCACCGTGGTCAGGCGGGCTGTATGAA 44
Db 1 ACTATAGGCCACCGTGGTCAGGCGGGCTGTGTTAAAAA 41

RESULT 11
US-09-424-283-5/C
; Sequences 5; Application US/09424283
; Patent No. 6437219
; GENERAL INFORMATION:
; APPLICANT: Grimes, et al.
; TITLE OF INVENTION: Sucrose binding proteins
; FILE REFERENCE: 4630-50206
; CURRENT FILING DATE: 1999-11-19
; PRIOR APPLICATION NUMBER: PCT/US98/10465
; PRIOR FILING DATE: 1998-05-21
; PRIOR APPLICATION NUMBER: US 60/047,568
; PRIOR FILING DATE: 1997-05-22
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 1924
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (119)..(1588)
US-09-424-283-5

Query Match 3.7% Score 37.8; DB 4; Length 1924;
Best Local Similarity 95.1%; Pred. No. 0.065; 0; Gaps 0;

Qy 1 CTTACTATAGGCCACGGTGGTCAGGCCGGCTGT 39
Db 1803 CTCACTATAGGCCACGGTGGTCAGGCCGGCTGT 1763

RESULT 12
US-09-913-014A-18
; Sequence 18; Application US/08913014A
; Patent No. 623878
; GENERAL INFORMATION:

;

APPLICANT: Nishi, Kazunori
; APPLICANT: Hikichi, Yukiko
; APPLICANT: Shintani, Yasushi
; TITLE OF INVENTION: NOVEL FAS LIGAND-LIKE PROTEIN, ITS
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David G. Conlin, Esq.
; STREET: 130 Water Street
; CITY: Boston,
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/913,014A
; FILING DATE: 04-SEP-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP97/02480
; FILING DATE: JULY 17, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: David G. Conlin
; REGISTRATION NUMBER: 27,026
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-523-3400
; TELEFAX: 617-523-6440
; TELEX:
; INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
; LENGTH: 48
; TYPE: Nucleic acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; MOLECULE TYPE: Synthetic DNA
US-08-913-014A-18

Query Match 3.6% Score 37.4; DB 4; Length 48;
Best Local Similarity 97.4%; Pred. No. 0.012; 0; Gaps 0;

Qy 1 CTTACTATAGGCCACGGTGGTCAGGCCGGCTGT 39
Db 10 CTCACTATAGGCCACGGTGGTCAGGCCGGCTGT 48

RESULT 13
US-09-570-367C-1/C
; Sequence 1; Application US/09570367C
; Patent No. 6338851
; GENERAL INFORMATION:
; APPLICANT: Gorczynski, Reginald M.
; TITLE OF INVENTION: Methods and Compositions for Immunomodulation
; FILE REFERENCE: 9579-21
; CURRENT APPLICATION NUMBER: US/09/570,367C
; CURRENT FILING DATE: 2000-05-05
; PRIOR APPLICATION NUMBER: 1997-11-07
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 2791
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-570-367C-1

Query Match 3.6% Score 37.4; DB 4; Length 2791;

RESULT 14 ; LENGTH: 1478
 Best Local Similarity 97.4%; Pred. No. 0.11;
 Matches 38; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 ; TYPE: DNA
 ; ORGANISM: Tagged Ctenocephalides felis
 US-09-545-814-30 ;
 Query Match 3.6%; Score 37.2; DB 4; Length 1478;
 Best Local Similarity 47.1%; Pred. No. 0.087;
 Matches 114; Conservative 0; Mismatches 128; Indels 0; Gaps 0;
 QY 66 ATATACTGCTGAGAGATAACATCACAAATTCAAGCTTCGATGTCACTACAGCCPACAGC 125
 DB 1023 AGACATTAATGGATAGCATTAATTCGAAAGTCTGGTACATGATTCCTTCCTG 1082
 QY 126 CATTCGGAGCAGGCCCTAGGACGGCTCAATAACTGGAGGAACGTGCCAAGATGTCG 185
 DB 1083 CAGCCTAAATTGGCAATGGGACAGGGATGGGATCTCCCGAGTTTGCCPTGTTGTT 1142
 QY 186 GATTACAAACAGTCTATCTGGGACACTCTAACGAGCTTGTATAGTCATTATATC 245
 DB 1143 GAAAGCCCCTGGCAATTCTGGACAAGTATAAAAATTGGCTTTGGAAAAAACTGCC 1202
 QY 246 TCCCCAACCGGCATTGTAAGCAGGCCAACGCAATTGCAATTGTTGTTGCTAC 305
 DB 1203 ACCTCTATCAGGACCCAGGATACTCCAAATCAGTCGAACCTGTAATTC 1262
 QY 306 TT 307
 DB 1263 GT 1264

Search completed: December 28, 2002, 01:26:51
 Job time : 61 secs

RESULT 15 ; LENGTH: 1478;
 Best Local Similarity 47.1%; Pred. No. 0.087;
 Matches 114; Conservative 0; Mismatches 128; Indels 0; Gaps 0;
 ; TYPE: DNA
 ; ORGANISM: Tagged Ctenocephalides felis
 US-09-545-814-28/C ;
 Sequence 28, Application US/09545814
 ; GENERAL INFORMATION:
 ; APPLICANT: Becher, Anna M.
 ; TITLE OF INVENTION: FLEA CHITINASE NUCLEIC ACID MOLECULES, PROTEINS AND
 FILE REFERENCE: FC-5-C1
 CURRENT APPLICATION NUMBER: US/09/545,814
 PRIORITY FILING DATE: 2000-04-07
 PRIORITY APPLICATION NUMBER: 60/128,833
 PRIORITY FILING DATE: 1999-04-09
 NUMBER OF SEQ ID NOS: 40
 SOFTWARE: Patentin Ver. 2.1
 SEQ ID NO 28
 LENGTH: 1478
 ; TYPE: DNA
 ; ORGANISM: Tagged Ctenocephalides felis
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)...(1467)
 US-09-545-814-28

Query Match 3.6%; Score 37.2; DB 4; Length 1478;
 Best Local Similarity 47.1%; Pred. No. 0.087;
 Matches 114; Conservative 0; Mismatches 128; Indels 0; Gaps 0;
 QY 66 ATATACTGCTGAGAGATAACATCACAAATTCAAGCTTCGATGTCACTACAGCCPACAGC 125
 DB 456 AGCATACGCTGATAGCATTAATTCGAAATGGTACATGATTCCTTCCTG 397
 QY 126 CATTCGGAGCAGGCCCTAGGACGGCTCAATAACTGGAGGAACCTGGCAAGATGTC 185
 DB 396 CAGCCTAAATTGGCAATGGGACACCCAGTGTATCCTCAGTTGGCTTTGGTTGTT 337
 QY 186 GATTACAAACAGTCTATCTGGGACAGTCTAACGCTTGTATAGTCATTATTC 245
 DB 336 GAAAGCCCCTGGCAATCTGGACAAGTATAAAAATTGGCTTTGGAAAAAACTGCC 277
 QY 246 TCCCCAACCGGCATTGTAAGCAGGCCAACGCAATTGCAATTGTTGCTAC 305
 DB 276 ACCTCTATCAGGACCCAGGATACTCCAAATCAGTCGAACCTGTAATTC 217
 QY 306 TT 307
 DB 216 GT 215

RESULT 15 ; LENGTH: 1478;
 US-09-545-814-30 ;
 Sequence 30, Application US/09545814
 ; Patent No. 641697
 ; GENERAL INFORMATION:
 ; APPLICANT: Becher, Anna M.
 ; TITLE OF INVENTION: FLEA CHITINASE NUCLEIC ACID MOLECULES, PROTEINS AND
 FILE REFERENCE: FC-5-C1
 CURRENT APPLICATION NUMBER: US/09/545,814
 PRIORITY FILING DATE: 2000-04-07
 PRIORITY APPLICATION NUMBER: 60/128,833
 PRIORITY FILING DATE: 1999-04-09
 NUMBER OF SEQ ID NOS: 40
 SOFTWARE: Patentin Ver. 2.1
 SEQ ID NO 30



OM nucleic - nucleic search, using sw model														
Run on: December 27, 2002, 23:38:28 ; Search time 252 Seconds (without alignments) 9204.596 Million cell updates/sec														
Title: US-09-945-376-3														
Score:	Perfect score:	1030	Sequence:	1 cttaatataggcaacgtcg.....tgatataataaggctctgg 1030	Scoring table:	IDENTITY_NUC	Gapext 1.0	Post-processing:	Minimum Match 0%					
Searched:	2185239 seqs, 1125999159 residues	Total number of hits satisfying chosen parameters:	4370478	Maximum DB seq length:	20000000000	Maximum Match 100%	Listing first 45 summaries	Database :	N_Geneseq_101002.*					
1: /SIDS2/gcgdata/geneseq/geneseq/geneseq -emb1/NA1980.DAT :*														
2: /SIDS2/gcgdata/geneseq/geneseq/geneseq -emb1/NA1981.DAT :*														
3: /SIDS2/gcgdata/geneseq/geneseq/geneseq -emb1/NA1982.DAT :*														
4: /SIDS2/gcgdata/geneseq/geneseq/geneseq -emb1/NA1983.DAT :*														
5: /SIDS2/gcgdata/geneseq/geneseq/geneseq -emb1/NA1984.DAT :*														
6: /SIDS2/gcgdata/geneseq/geneseq/geneseq -emb1/NA1985.DAT :*														
7: /SIDS2/gcgdata/geneseq/geneseq/geneseq -emb1/NA1986.DAT :*														
8: /SIDS2/gcgdata/geneseq/geneseq/geneseq -emb1/NA1987.DAT :*														
9: /SIDS2/gcgdata/geneseq/geneseq/geneseq -emb1/NA1988.DAT :*														
10: /SIDS2/gcgdata/geneseq/geneseq/geneseq -emb1/NA1989.DAT :*														
11: /SIDS2/gcgdata/geneseq/geneseq/geneseq -emb1/NA1990.DAT :*														
12: /SIDS2/gcgdata/geneseq/geneseq/geneseq -emb1/NA1991.DAT :*														
13: /SIDS2/gcgdata/geneseq/geneseq/geneseq -emb1/NA1992.DAT :*														
14: /SIDS2/gcgdata/geneseq/geneseq/geneseq -emb1/NA1993.DAT :*														
15: /SIDS2/gcgdata/geneseq/geneseq/geneseq -emb1/NA1994.DAT :*														
16: /SIDS2/gcgdata/geneseq/geneseq/geneseq -emb1/NA1995.DAT :*														
17: /SIDS2/gcgdata/geneseq/geneseq/geneseq -emb1/NA1996.DAT :*														
18: /SIDS2/gcgdata/geneseq/geneseq/geneseq -emb1/NA1997.DAT :*														
19: /SIDS2/gcgdata/geneseq/geneseq/geneseq -emb1/NA1998.DAT :*														
20: /SIDS2/gcgdata/geneseq/geneseq/geneseq -emb1/NA1999.DAT :*														
21: /SIDS2/gcgdata/geneseq/geneseq/geneseq -emb1/NA2000.DAT :*														
22: /SIDS2/gcgdata/geneseq/geneseq/geneseq -emb1/NA2001A.DAT :*														
23: /SIDS2/gcgdata/geneseq/geneseq/geneseq -emb1/NA2001B.DAT :*														
24: /SIDS2/gcgdata/geneseq/geneseq/geneseq -emb1/NA2002.DAT :*														
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.														
SUMMARIES														
Result No.	Score	Query Match	Length	DB	ID	Description	8	Query	%					
C 1	4.5	4.4	3548	24	AAD38882	Alfalfa plastocyanin Karp-1 DNA. Cotton promoter region sequence.	XX	XX	XX					
C 2	42.4	4.1	6078	22	AAF31861	Rat glyceraldehyde-3-phosphate dehydrogenase gene promoter.	XX	XX	XX					
C 3	41	4.0	2614	19	AAV40521	Rat glyceraldehyde-3-phosphate dehydrogenase gene promoter.	XX	XX	XX					
C 4	40.6	3.9	1038	21	AAC62796	Pinus radiata promoter.	XX	XX	XX					
C 5	40.6	3.9	1038	24	ABK17061	Rice another specific O-methyl transferase gene.	XX	XX	XX					
C 6	40.4	3.9	598	24	AAS15016	Eucalyptus grandis Arabidopsis BONSAI	PT	PT	PT					
C 7	40.4	3.9	2096	21	AAC62810	Oligonucleotide primers that amplify sequences upstream or downstream of the target sequence.	PT	PT	PT					
C 8	40.4	3.9	2096	24	ABA91705	Isolating and characterizing an expression regulatory sequence for expressing recombinant polypeptides and/or RNA, comprises producing	PT	PT	PT					
C 9	40.4	3.9	2611	24	ABA91255	OM nucleic acid primers that amplify sequences upstream or downstream of the target sequence.	PT	PT	PT					

PT of cDNAs -
 XX Example 2; Page 68-69; 74pp; English.

CC The invention relates to a method for isolating and characterising an expression regulatory sequence for the expression of recombinant polypeptides and/or RNA. The method comprising producing at least one oligonucleotide primer from cDNA library, where the oligonucleotide primer allows amplification of genomic sequences upstream or downstream of the cDNAs. The method is useful for isolating, characterising and identifying a large number of known and unknown promoters that are active under any desired environmental condition to which a cell may be exposed and not just to the exemplified isolation of promoters that are capable of expressing in specific conditions. The methods are also useful for cloning genes from any host, or from a specific tissue with such host, from which a cDNA library may be constructed; for the identification and isolation of analogous promoters, signal peptides and structural genes in several species of multicellular and unicellular organisms and as a high throughput identification system of candidate therapeutic targets. The promoter sequences may be used to regulate the synthesis of polypeptides. The present sequence is alfalfa plastocyanin gene coding fragment.

XX Sequence 3548 BP; 1112 A; 576 C; 727 G; 1133 T; 0 other;
 SQ Query Match 4.4%; Score 45; DB 24; Length 3548;
 Best Local Similarity 90.6%; Pred. No. 0.0029; Indels 0; Gaps 0;
 Matches 48; Conservative 0; Mismatches 5;

QY 2 TTACTATAGGCCACGGCTGGTGGTGGACGCCCGGCGTGTATGAAAGTTGGAACCC 54
 Db 3529 TTACTATAGGCCACGGCTGGTGGTGGACGCCCGGCGTGTATGAAAGTTGGAACCC 3477

RESULT 2
 AAF31861
 ID AAF31861 standard; DNA; 6078 BP.
 XX AC AAF31861;
 XX DT 12-APR-2001 (first entry)
 DE Human KARP-1 DNA.
 XX Human; KARP-1; Ku86 autoantigen related protein; cancer; immune disorder; biliary tract cancer; leucine zipper protein; cytosolic; immunosuppressant; gene therapy; KARP-1 inhibitor; ds.
 OS Homo sapiens.
 PN US6171857-B1.
 XX PD 09-JAN-2001.
 XX PF 16-OCT-1998; 98US-0173914.
 XX PR 17-OCT-1997; 97US-0064557.
 PA (UYBR-) UNIV BROWN RES FOUND.
 PI Hendrickson EA;
 XX WPI; 2001-146208/15.
 DR P-PSDB; AA666590.
 XX Novel nucleic acids encoding leucine zipper protein, KARP-1 polypeptide, useful for treating cancer and immune deficiency disorder

CC The present sequence is given in a specification relating to an isolated Ku86 Autoantigen Related Protein (KARP-1) nucleic acid molecule. The

CC KARP-1 nucleic acid and KARP-1 protein are useful for the treatment and/or diagnosis of diseases such as cancer and immune deficiency disorders. They are useful in combination with a KARP-1 inhibitor that inhibits double stranded DNA base repair. Inhibitors of KARP-1 are useful in the diagnosis or treatment of conditions characterised by the loss of KARP-1 activity and in the treatment of cancer, e.g. biliary tract cancer.

XX Sequence 6078 BP; 1806 A; 1301 C; 1392 G; 1543 T; 36 other;
 SQ Query Match 4.1%; Score 42.4; DB 22; Length 6078;
 Best Local Similarity 97.7%; Pred. No. 0.028; Indels 0; Gaps 0;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTTACTATAGGCCACGGCTGGTGGACGCCCGGCGTGTATGAA 44
 Db 9 CTTACTATAGGCCACGGCTGGTGGACGCCCGGCGTGTATGAA 52

RESULT 3
 AAV40621
 ID AAV40621 standard; DNA; 2614 BP.
 XX AC AAV40621;
 XX DT 26-OCT-1998 (first entry)
 DE Cotton promoter region from an expansin gene.
 XX KW Cotton fibre expansin Promoter; transformation; transgenic; ss.
 XX OS Gossypium hirsutum
 PN WO9830698-A1.
 XX PR 07-JAN-1997; 97US-0034914.
 XX PA (CALJ) CALGENE INC.
 PI Pear JR, Stalker DM;
 XX DR WPI; 1998-390143/34.
 XX PT New DNA containing the promoter of the cotton expansin gene - used for tissue-selective expression of genes that alter cotton fibre phenotype
 XX PS Claim 2; Fig 1a-d; 26pp; English.

Query Match 4.0%; Score 41; DB 19; Length 2614;
 Best Local Similarity 10.0%; Pred. No. 0.048;
 Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTTACTATAGGCCACGGCTGGTGGACGCCCGGCGTGTATGAA 41
 Db 83 CTTACTATAGGCCACGGCTGGTGGACGCCCGGCGTGTAT 123

RESULT 4
 AAC62796

ID AAC62796 standard; DNA; 1038 BP.
 XX AAC62796;
 AC XX
 XX DT 02-FEB-2001 (first entry)
 XX DE Senescence-like protein promoter coding sequence #1.
 XX KW Promoter; eucalyptus; pine; gene transcription; ds.
 OS XX
 Pinus radiata.
 XX PN WO200058474-A1.
 XX PD 05-OCT-2000.
 XX PF 24-FEB-2000; 2000WO-NZ000018.
 XX PR 25-MAR-1999; 990US-0276599.
 PR 30-JUL-1999; 990US-0146591.
 XX PA (GENE-) GENESIS RES & DEV CORP LTD.
 PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
 XX PI Perera R, Rice SJ, Eagleton CK;
 XX WPI; 2000-647236/62.
 DR P-PSDB; AAB28136.
 XX PT Novel promoter sequences useful for modulating transcription of plant
 DNA sequences of interest and production of polypeptides -
 XX PS Claim 1; Pages 57-58; 93pp; English.
 XX The present invention relates to promoter sequences from eucalyptus
 and pine. The present sequence is one such promoter. This sequence is
 useful for modulating the transcription of DNA sequences of interest. The
 sequences may also be used to tag or identify an organism or its
 reproductive material.
 XX SQ Sequence 1038 BP; 298 A; 220 C; 233 G; 287 T; 0 other;
 SQ Score 40.6; DB 21; Length 1038;
 Query Match 3.9%; Best Local Similarity 71.1%; Pred. No. 0.037;
 Best Local Similarity 71.1%; Pred. No. 0.037;
 Matches 69; Conservative 0; Mismatches 24; Indels 4; Gaps 1;
 QY 2 TTACTATAGGGCACCGCTGCTCGACGGCCGGCTGGTATGAAGCTGGAACTCAGTG 61
 Db 4 TTACTATAGGGCACCGCTGCTCGACGGCCGGCTGGTATGAAGCTGGAACTCAAAG 59
 QY 62 ATGGCATATAGTCAGAGATAAACATCACAACTACA 98
 Db 60 CTGTGTAAATTCAATCTCTGTTACCATACAACTTCTCA 96

RESULT 5
 ABK17061 ID ABK17061 standard; cDNA; 1038 BP.
 XX AC ABK17061;
 XX DT 26-MAR-2002 (first entry)
 XX DE Pinus radiata promoter polynucleotide #20.
 XX KW Promoter; pine; leaf; root; flower; pollen; bud; meristem; xylogenesis;
 KW temporally regulated promoter; Pinus radiata; Eucalyptus grandis; ss;
 KW PCR primer.
 XX OS Pinus radiata.
 XX PN WO200198485-A1.
 XX

PD 27-DEC-2001.
 XX PF 20-JUN-2001; 2001WO-NZ000115.
 XX PR 20-JUN-2000; 2000US-0598401.
 PR 28-NOV-2000; 2000US-0724624.
 XX PA (GENE-) GENESIS RES & DEV CORP LTD.
 PA (FLET-) FLETCHER CHALLENGE FORESTS IND LTD.
 XX WPI; 2002-114582/15.
 PI Perera R, Rice S, Eagleton C, Lasham A;
 XX DR P-PSDB; AAU80734.

PD Novel polynucleotide promoter sequences from Pine and Eucalyptus useful
 PT for modifying expression of endogenous and/or heterologous
 PT polynucleotides in transgenic plants -
 XX PS Claim 1; Page 73-74; 121pp; English.
 XX The invention relates to isolated promoter sequences from Pinus radiata
 CC and Eucalyptus grandis, comprising a sequence chosen from leaf-, root-,
 CC flower-, pollen-, bud-, meristem-specific promoters or temporally
 CC regulated promoters such as xylogenesis-specific promoters. The promoter
 CC polypeptides and their related polynucleotides are useful in the
 CC production of genetic constructs used for modifying gene expression in a
 CC target organism, in particular a plant. The method is useful for
 CC modifying expression of a polynucleotide that comprises an intron
 CC sequence, through removal of the intron sequence. The method is useful
 CC for modifying growth and development of plants, and cellular responses to
 CC external stimulus, such as environmental factors and disease pathogens.
 CC The sequences are useful in genome and physical mapping, in positional
 CC cloning of genes, in various assays to determine biological activity, to
 CC raise antibodies, to isolate corresponding interacting proteins and other
 CC compounds and to quantitatively determine levels of interacting proteins
 CC or other compounds. Sequences ABK17016-ABK17125 represent Pinus radiata
 CC and Eucalyptus grandis polynucleotides and PCR primers used in the method
 CC of the invention.
 XX SQ Sequence 1038 BP; 298 A; 220 C; 233 G; 287 T; 0 other;
 SQ Score 40.6; DB 24; Length 1038;
 Query Match 3.9%; Best Local Similarity 71.1%; Pred. No. 0.037;
 Best Local Similarity 71.1%; Pred. No. 0.037;
 Matches 69; Conservative 0; Mismatches 24; Indels 4; Gaps 1;
 QY 2 TTACTATAGGGCACCGCTGCTCGACGGCCGGCTGGTATGAAGCTGGAACTCAGTG 61
 Db 4 TTACTATAGGGCACCGCTGCTCGACGGCCGGCTGGTATGAAGCTGGAACTCAAAG 59
 QY 62 ATGGCATATAGTCAGAGATAAACATCACAACTACA 98
 Db 60 CTGTGTAAATTCAATCTCTGTTACCATACAACTTCTCA 96

RESULT 6
 AAS15016 ID AAS15016 standard; cDNA; 598 BP.
 XX AC AAS15016;
 XX DT 14-FEB-2002 (first entry)
 XX DE Rice anther-specific promoter #2.
 XX KW Rice; promoter; gene therapy; transcription; monocot; anther;
 KW chromosome mapping; gene mapping; antisense technology;
 KW plant genetic engineering; ss.
 XX OS Oryza sativa.
 XX PN WO200181606-A2.
 XX

page 4

CC production of genetic constructs, used for modifying gene expression in a target organism, in particular a plant. The method is useful for CC modifying expression of a polynucleotide that comprises an intron CC sequence, through removal of the intron sequence. The method is useful CC for modifying growth and development of plants, and cellular responses to external stimulus, such as environmental factors and disease pathogens. CC The sequences are useful in genome and physical mapping, in positional CC cloning of genes, in various assays to determine biological activity, to CC raise antibodies, to isolate corresponding proteins and other CC compounds, and to quantitatively determine levels of interacting proteins CC or other compounds. Sequences ABK17016-ABK17125 represent *Pinus radiata* CC and *Eucalyptus grandis* polynucleotides and PCR primers used in the method CC of the invention.

XX Sequence 2096 BP; 537 A; 535 C; 539 G; 485 T; 0 other;
 Query Match 3.9%; Score 40.4; DB 24; Length 2096;
 Best Local Similarity 97.6%; Pred. No. 0.065;
 Matches 41; Conservative 0; Mismatches 1; Indels 0; gaps 0;
 Qy 2 TTACTATAGGCCACGGCTGGTGTGACGGCCGGCTGGTATGA 43
 Db 3 TTACTATAGGCCACGGCTGGTGTGACGGCCGGCTGGTCTGA 44

RESULT 9

ABA91255/c
 ID ABA91255 standard; DNA; 2611 BP.
 XX AC ABA91255;
 XX DT 04-APR-2002 (first entry)
 XX DE Arabidopsis BONSAII (BON1) gene, involved in growth homeostasis.
 XX KW BONSAII; BON1; phospholipid binding protein; growth; homeostasis;
 KW thermotolerance; transgenic plant; plant; gene; ds.
 XX OS Arabidopsis thaliana.
 XX PN WO200200697-A2.
 XX PD 03-JAN-2002.
 XX PF 25-JUN-2001; 2001WO-US200172.
 XX PR 23-JUN-2000; 2000US-213863P.
 XX PA (WHED) WHITEHEAD INST BIOMEDICAL RES.
 XX PI Hua J, Griseai P, Fink GR;
 XX WPI; 2002-139999/18.
 DR

XX New phospholipid binding proteins and nucleic acids, useful for PT modulating plant growth homeostasis, controlling cell expansion and PT cell division, or producing plants where larger fruits and increased PT biomass are desired -
 PS Disclosure; Fig 8A; 78pp; English.
 XX The present sequence is that of the BONSAII (BON1) gene of CC Arabidopsis thaliana. The coding region of the BON1 gene is CC given in ABA91255. The BON1 gene permits wild-type Arabidopsis CC plants to maintain a relatively constant size over a wide range of CC temperatures. Thus, bon1 null mutants produce miniature fertile C. plants at 22 degrees C, but a wild-type phenotype at 28 degrees C. CC BON1 has a direct role in regulating cell expansion and cell CC division at temperatures lower than those at which Arabidopsis is CC normally grown. The BON1 protein (see AAM50644) contains a CC Ca²⁺-dependent phospholipid binding domain and is associated with CC the plasma membrane. BON1 belongs to the copine gene family, which CC is conserved from protozoa to humans. The invention is directed to CC

CC isolated BON1, BON2, BON3, BAP1 and BAL nucleic acids (see CC ABA91256-00), which encode proteins (see AAM50644-48) that are CC necessary for normal growth, controlling cell expansion and cell CC division, affecting the size and rate at which the plant grows when CC exposed to lower temperatures. Transgenic plants are provided that CC are smaller than the wild-type as a result of inhibition of BON1, CC BON2, BON3, BAP1 and/or BAL, especially angiosperms and gymnosperms, CC ornamental plants and turfgrass. Transgenic plants are also CC provided that are larger than the wild-type as a result of enhancement of BON1, BON2, BON3, BAP1 and/or BAL, especially crop CC plants and biomass plants. Modulation of these genes provides CC increased yield, or growth at a higher altitude or lower CC temperature.
 XX SQ Sequence 2611 BP; 809 A; 429 C; 498 G; 866 T; 9 other;

Query Match 3.9%; Score 40.4; DB 24; Length 2611;
 Best Local Similarity 97.6%; Pred. No. 0.074;
 Matches 41; Conservative 0; Mismatches 1; Indels 0; gaps 0;
 Qy 1 CTTACTATAGGCCACGGCTGGTGTGACGGCCGGCTGGTATG 42
 Db 2603 CTTACTATAGGCCACCGTGTGACGGCCGGCTGGTCTG 2562

RESULT 10

AAC62780
 ID AAC62780 standard; DNA; 411 BP.
 XX AC AAC62780;
 XX DT 02-FEB-2001 (first entry)
 XX DE Flower specific promoter coding sequence #2.
 XX KW Promoter; eucalyptus; pine; gene transcription; ds.
 XX OS Eucalyptus grandis.
 XX PN WO200058474-A1.
 XX PD 05-OCT-2000.
 XX PF 24-FEB-2000; 2000WO-NZ00018.
 XX PR 25-MAR-1999; 99US-0276599.
 XX PR 30-JUL-1999; 99US-0146591.
 XX PA (GENE-) GENESIS RES & DEV CORP LTD.
 PA (FLET-) FLETCHER CHALLENGE FORBSTS LTD.
 XX PI Perera R, Rice SJ, Eagleton CK;
 XX DR WPI; 2000-647236/62.

XX The present invention relates to promoter sequences from eucalyptus CC and pine. The present sequence is one such promoter. This sequence is CC useful for modulating the transcription of DNA sequences of interest. The CC sequences may also be used to tag or identify an organism or its CC reproductive material.
 XX SQ Sequence 411 BP; 112 A; 118 C; 80 G; 101 T; 0 other;
 XX PS Claim 1; Page 53; 93pp; English.
 XX The present invention relates to promoter sequences from eucalyptus CC and pine. The present sequence is one such promoter. This sequence is CC useful for modulating the transcription of DNA sequences of interest. The CC sequences may also be used to tag or identify an organism or its CC reproductive material.
 XX SQ Sequence 411 BP; 112 A; 118 C; 80 G; 101 T; 0 other;
 XX PS The present invention relates to promoter sequences from eucalyptus CC and pine. The present sequence is one such promoter. This sequence is CC useful for modulating the transcription of DNA sequences of interest. The CC sequences may also be used to tag or identify an organism or its CC reproductive material.
 XX SQ Sequence 411 BP; 112 A; 118 C; 80 G; 101 T; 0 other;
 XX PS Query Match 3.9%; Score 40.2; DB 21; Length 411;
 XX Best Local Similarity 93.3%; Pred. No. 0.029;
 XX Matches 42; Conservative 0; Mismatches 3; Indels 0; gaps 0;
 XX Qy 4 ACTATAGGGCACGGTGTGACGGCCGGCTGGTATGAACTG 48

Db 1 ACTATAGGGCACGGTGGTGACGCCGGCTGTCGAAGCTG 45
RESULT 11
 ABK17045 standard; cDNA; 411 BP.
 XX
 AC ABK17045;
 XX
 DT 26-MAR-2002 (first entry)
 DE Eucalyptus grandis promoter polynucleotide #16.
 KW Promoter; pine; leaf; flower; pollen; bud; meristem; xylogenesis; temporally regulated promoter; Pinus radiata; Eucalyptus grandis; ss; PCR primer.
 XX
 OS Eucalyptus grandis.
 XX
 PN WO200198485-A1.
 XX
 PD 27-DEC-2001.
 XX
 PF 20-JUN-2001; 2001WO-NZ00115.
 XX
 PR 20-JUN-2000; 2000US-0598401.
 PR 28-NOV-2000; 2000US-0724624.
 XX
 PA (GENE-) GENESIS RES & DEV CORP LTD.
 PA (FLET-) FLETCHER CHALLENGE FORESTS IND LTD.
 XX
 PI Perera R, Rice S, Easleton C, Lasham A;
 XX
 DR 2002-114583/15.
 XX
 PS Claim 1; Page 68-69; 121pp; English.
 XX
 CC The invention relates to isolated promoter sequences from *Pinus radiata* and *Eucalyptus grandis*, comprising a sequence chosen from leaf-, root-, flower-, pollen-, bud-, meristem specific promoters or temporally regulated promoters such as xylogenesis-specific promoters. The promoter polypeptides and their related polynucleotides are useful in the production of genetic constructs, used for modifying gene expression in a target organism, in particular a plant. The method is useful for modifying expression of a polynucleotide that comprises an intron sequence, through removal of the intron sequence. The method is useful for modifying growth and development of plants, and cellular responses to external stimuli, such as environmental factors and disease pathogens. The sequences are useful in genome and physical mapping, in positional cloning of genes, in various assays to determine biological activity, to raise antibodies, to isolate corresponding interacting proteins and other compounds, and to quantitatively determine levels of interacting proteins or other compounds. Sequences ABK17016-ABK1725 represent *Pinus radiata* and *Eucalyptus grandis* polynucleotides and PCR primers used in the method of the invention.
 XX
 Sequence 411 BP; 112 A; 118 C; 80 G; 101 T; 0 other;
 SO Query Match 3.9%; Score 40.2; DB 24; Length 411;
 Best Local Similarity 93.3%; Pred. No. 0.029; Mismatches 3; Indels 0; Gaps 0;
 XX
 Qy 4 ACTATAGGGCACGGTGGTGACGCCGGCTGTCGAAGCTG 48
 Best Local Similarity 93.3%; Pred. No. 0.029; Mismatches 3; Indels 0; Gaps 0;
 Db 1 ACTATAGGGCACGGTGGTGACGCCGGCTGTCGAAGCTG 45
 RESULT 12
 AAA5054
 ID AAA5054 standard; cDNA; 594 BP.
 XX
 AC AAA75054;
 XX

AXX35652 standard; cDNA; 594 BP.
 XX
 AC AXX35652;
 XX
 DT 09-JUL-1999 (first entry)
 DE 5' region of human heparanase cDNA.
 XX
 KW Heparanase; hpa; modulator; heparin-binding growth factor; cellular response; cytokine; cell interaction; plasma lipoprotein; cellular susceptibility; infection; disintegration; neurodegenerative plaque; wound healing; angiogenesis; restenosis; atherosclerosis; inflammation; neurodegenerative disease; neutralise; plasma heparin; micrometastasis; autoimmune lesion; renal failure; ss.
 KW
 XX
 OS Homo sapiens.
 XX
 PN WO9911798-A1.
 XX
 PD 11-MAR-1999.
 XX
 PF 31-AUG-1998; 98WO-US17954.
 XX
 PR 02-JUL-1998; 98US-0109386.
 PR 02-SEP-1997; 97US-0922170.
 XX
 PA (FRIE-) FRIEDMAN M M.
 PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
 PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.
 XX
 PI Feinstein E, Pecker I, Vlodavsky I;
 XX
 DR 1999-302255/25.
 XX
 PT New human polynucleotide useful for treating angiogenesis, restenosis, and inflammation
 XX
 PS Example 8; Page 69; 63pp; English.
 XX
 CC The specification describes a polypeptide having heparanase (hpa) activity. The recombinant protein is used as a modulator of heparin-binding growth factors, cellular responses to heparin-binding growth factors and cytokines, cell interaction with plasma lipoproteins, cellular susceptibility to viral, protozoal and bacterial infections or disintegration of neurodegenerative plaques. Heparanase may be useful for conditions such as wound healing, angiogenesis, restenosis, atherosclerosis, inflammation, neurodegenerative diseases, and viral infections. Mammalian heparanase can be used to neutralize plasma heparin, and anti-heparanase antibodies may be applied for immunodetection and diagnosis of micrometastases, autoimmune lesions, and renal failure in biopsy specimens, plasma samples, and body fluids. The present sequence represents the 5' sequence of human heparanase cDNA.
 XX
 Sequence 594 BP; 133 A; 134 C; 209 G; 118 T; 0 other;
 CC Query Match 3.9%; Score 40; DB 20; Length 594;
 CC Best Local Similarity 100.0%; Pred. No. 0.041;
 CC Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 2 TTACTATAGGGCACGGTGGTGACGCCGGCTGTCGAAGCTG 41
 Db 2 TTACTATAGGGCACGGTGGTGACGCCGGCTGTCGAAGCTG 41
 RESULT 13
 AAA5054
 ID AAA5054 standard; cDNA; 594 BP.
 XX
 AC AAA75054;
 XX

DT 15-JAN-2001 (first entry)
 XX Upstream sequence of human cDNA encoding heparanase.
 DE Human; heparanase; gene therapy; tumour; inflammation; autoimmunity;
 XX heparin-binding growth factor; cytokine; neurodegenerative plaque;
 KW wound healing; infection; burn; angiogenesis; restenosis;
 KW atherosclerosis; inflammation; neurodegenerative disease;
 KW Gerstmann-Straussler Syndrome; Creutzfeldt-Jakob disease; ds.
 XX Homo sapiens.
 OS WO200053178-A1.
 XX PD 08-SEP-2000.
 XX PF 14-FEB-2000; 2000WO-US03542.
 XX PR 01-MAR-1999; 99US-0258892.
 XX PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.
 PA (HADA-) HADIST MEDICAL RES SERVICES & DEV.
 PA (PRIE-) FRIEDMAN M M.
 XX PI Pecker I, Vladavsky I, Feinstein E;
 XX DR WPI; 2000-579289/54.
 XX New polynucleotides encoding a polypeptide having heparanase activity,
 PT useful in wound healing and in gene therapy, particularly in treating
 PT tumour, inflammation, autoimmunity, neurodegenerative diseases -
 XX Example 7; Page 127; 152pp; English.
 XX The present sequence is an upstream sequence of cDNA encoding encodes a
 CC human protein with heparanase catalytic activity. The heparanase (hpa)
 CC polynucleotide is useful in gene therapy, particularly in treating (hpa)
 CC tumour, inflammation or autoimmunity. Particularly, the polynucleotide
 CC is useful in modulating the bioavailability of heparin-binding growth
 CC factors, cellular responses to heparin-binding growth factors (e.g. bFGF)
 CC and cytokines (e.g. interleukin (IL)-8), cell interaction with plasma
 CC lipoproteins, cellular susceptibility to certain viral and some bacterial
 CC and protozoa infections, or disintegration of neurodegenerative plaques.
 CC The polynucleotide is also useful in wound healing (e.g. thermal,
 CC chemical or radiation burns), and in the treatment of angiogenesis,
 CC restenosis, atherosclerosis, inflammation, neurodegenerative diseases
 CC (Gerstmann-Straussler Syndrome or Creutzfeldt-Jakob disease), and some
 CC viral, bacterial or protozoa infections.
 XX Sequence 594 BP; 133 A; 134 C; 209 G; 118 T; 0 other;
 XX Query Match 3.9%; score 40; DB 21; Length 594;
 Best Local Similarity 100.0%; Pred. No. 0.041;
 Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 PT XX OS Homo sapiens.
 PN XX JP10295383-A.
 PD XX 10-NOV-1998.
 PP XX 24-APR-1997; 97JP-0107624.
 PR XX 24-APR-1997; 97JP-0107624.
 PA XX (ADS-K-) ADVANCED SKIN RES KENKYUSHO KK.
 WPI XX 1999-038280/04.
 DR XX A promoter for hyaluronate synthase gene - used for screening of
 PT modulators of the enzyme
 PT XX Claim 1; Page 5-6; 7pp; Japanese.
 PS XX This sequence represents the DNA of the invention, and can act as a
 CC promoter for the hyaluronate synthase gene. It was isolated from human
 CC CDNA sources. The DNA is useful for screening a drug enhancing or
 CC inhibiting production of hyaluronic acid.
 XX SQ Sequence 1680 BP; 452 A; 381 C; 467 G; 380 T; 0 other;
 XX Query Match 3.9%; score 40; DB 20; Length 1680;
 Best Local Similarity 100.0%; Pred. No. 0.077;
 Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 PT XX RESULT 15
 AAZ51547 standard; DNA; 1224 BP.
 ID AAZ51547
 XX AC AAZ51547;
 XX DT 21-JUN-2000 (first entry)
 DE XX Seed-preferred promoter-2 from maize end gene.
 XX KW Seed-preferred promoter; maize; end gene; endosperm; end1; end2;
 KW fatty acid; starch profile; carbohydrate profile; amino acid content;
 KW agronomic; insect resistance; disease resistance; herbicide resistance;
 KW grain characteristic; ds.
 XX OS Zea mays.
 XX WO200012733-A1.
 PN XX 09-MAR-2000.
 PD XX 25-AUG-1990;
 PR XX 28-AUG-1998;
 PA XX (PION-) PIONEER HI-BRED INT INC.
 PI XX Martino-Catt SJ, LaPpegard KK, Olsen O, Linnestad C, Abbott SE;
 XX DR WPI; 2000-256648/22.
 XX PT Maize promoter driving transcription in a seed-preferred manner, for
 PT stably transforming plant cells -
 XX PS Claim 1; Pages 35-36; 43pp; English.
 XX CC The present sequence is a seed-preferred promoter isolated from maize
 CC endosperm specific genes, end1 or end2. The promoter is capable of

CC driving transcription in a seed-preferred manner and can be used in an
CC expression cassette, to stably transform plant cells. The expression
CC cassette can be used to modify the fatty acid content of seeds, alter
CC the starch or carbohydrate profile, and/or alter the amino acid content
CC of the seed. It can also be used to deliver genes encoding important
CC traits for agronomics, including insect resistance, disease resistance,
CC herbicide resistance, and grain characteristics.

XX SQ Sequence 1224 BP; 378 A; 239 C; 182 G; 425 T; 0 other;

Query Match 3.98; Score 39.8; DB 21; Length 1224;
Best Local Similarity 95.3%; Pred. No. 0.074;
Matches 41; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 3 TACTATAGGCACGGCAGCGTGGTGGACGGGCCGGCTGGTATGAAAG 45
Db 1 TACTATAGGCACGGCAGCGTGGTGGACGGGCCGGCTGGTAAAG 43

Search completed: December 28, 2002, 00:07:43
Job time : 254 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

8							
	Result No.	Score	Query Match Length	DB	ID	Description	
OM nucleic - nucleic search, using sw model	c 1	49.8	4.8	125020	9	AF429315	Homo sapi
Run on: December 27, 2002, 19:35:28 ; Search time 2673 Seconds	c 2	44.8	4.3	125020	9	AF429315	Homo sapi
(without alignments)	c 3	44.4	4.3	2167878	2	AL773534	Mus muscu
11214.325 Million cell updates/sec	c 4	42.8	4.2	685	8	AF457660	Castanea
Title: US-09-945-376-3	c 5	42.4	4.1	1423	9	AF039526	Homo sapi
Perfect score: 1030	c 6	42.4	4.1	6078	6	AR124194	Sequence
Sequence: 1 cttactataggccgcgtg.....tgtataataacaaggctgg 1030	c 7	41.8	4.1	92633	3	PFM4AP1_3	Continuation (4 of
Scoring table: IDENTITY_NUC	c 8	41.8	4.1	11000	2	PFM4AP1_2	Continuation (3 of
Gapop 10.0 , Gapext 1.0	c 9	41	4.0	4260	5	AY124482	Danio rer
AFO51769 Homo sapi	c 10	41	4.0	11957	9	AF051769	AY124482
Searched: 2054640 seqs, 14551402878 residues	c 11	40.8	4.0	1094	8	AF075270	Hordeum v
Total number of hits satisfying chosen parameters:	c 12	40.8	4.0	166703	9	AL139378	Human DNA
4109280	c 13	40.6	3.9	156325	9	HS13498	AL031655
Minimum DB seq length: 0	c 14	40.4	3.9	598	6	AX320057	Human DNA
Maximum DB seq length: 20000000000	c 15	40.4	3.9	1463	3	AF153014	Trichomon
Post-processing: Minimum Match 0%	c 16	40.4	3.9	2611	6	AX380786	Sequence
Listing first 45 summaries	c 17	40.2	3.9	2463	9	AB021922	Homo sapi
Maximum Match 100%	c 18	40.2	3.9	22031	9	AC025822	Homo sapi
Database : GenEmbl:*	c 19	40	3.9	498	10	RNU53907	Rattus norv
1: gb_ba:*	c 20	40	3.9	951	3	AF233737	Agrotis i
2: gb_htg:*	c 21	40	3.9	985	6	AX254342	Sequence
3: gb_in:*	c 22	40	3.9	1499	6	AX253428	Sequence
4: gb_lm:*	c 23	40	3.9	1958	10	AF114032	Mus muscu
5: gb_ov:*	c 24	39.8	3.9	6805	6	AF289605	Mus muscu
6: gb_pat:*	c 25	39.8	3.9	2385	6	AX041981	Sequence
7: gb_ph:*	c 26	39.8	3.9	34980	6	AX344560	Sequence
8: gb_pr:*	c 27	39.6	3.8	7218	6	I66494	Sequence
9: gb_pr:*	c 28	39.4	3.8	640	10	U63199	Mus muscu
10: gb_ro:*	c 29	39.4	3.8	1135	9	AF074397	Homo sapi
11: gb_sts:*	c 30	39.4	3.8	2492	6	AX098431	Sequence
12: qb_sy:*	c 31	39.4	3.8	2867	6	E26792	Novel prote
13: qb_un:*	c 32	39.4	3.8	4185	1	AF237414	Ehrlichia
14: qb_vt:*	c 33	39	3.8	516	6	AX320056	Sequence
15: em_ba:*	c 34	39	3.8	615	10	AF077743	Sequence
16: em_fun:*	c 35	39	3.8	2184	6	AX233431	Sequence
17: em_hum:*	c 36	39	3.8	2785	10	AF162890	Mus muscu
18: em_in:*	c 37	38.8	3.8	104526	9	AC087887	Homo sapi
19: em_mu:*	c 38	38.8	3.8	149440	2	AC024319	Homo sapi
20: em_om:*	c 39	38.8	3.8	196049	2	AC105841	Rattus no
21: em_or:*	c 40	38.8	3.8	29530	1	AL591786	Sinorhizo
22: em_ov:*	c 41	38.8	3.8	329100	1	SME591787	Castanea
23: em_ph:*	c 42	38.6	3.7	1157	8	AF457661	Sequence
24: em_p1:*	c 43	38.6	3.7	5031	6	AX085162	Sequence
25: em_p1:*	c 44	38.6	3.7	5031	6	AX085359	Mus muscu
26: em_ro:*	c 45	38.6	3.7	180968	2	AC078996	Sequence
27: em_sts:*							
28: em_un:*							
29: em_vt:*							
30: em_htg_hum:*							
31: em_htg_inv:*							
32: em_htg_other:*							
33: em_htg_mus:*							
34: em_htg_pln:*							
35: em_htg_rnd:*							
36: em_htg_mam:*							
37: em_htg_vrt:*							
38: em_sy:*							
39: em_htgo_hum:*							
40: em_htgo_mus:*							
41: em_htgo_other:*							

ALIGNMENTS

RESULT 1	AF429315/c	LOCUS AF429315	125020 bp	DNA linear	PRI 18-JAN-2002
DEFINITION Homo sapiens junctophilin 3 (JPH3)	gene, partial cds.	ACCESSION AF429315			
VERSION 1	GI:17646244	VERSION 1			
KEYWORDS		KEYWORDS			
SOURCE		Homo sapiens			
ORGANISM		Hominoidea; Eutheria; Primates; Catarrhini; Homidae; Homo.			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
REFERENCE 1 (bases 1 to 125020)					
AUTHORS Ingarsoll-Ashworth,R.G., Fleisher,A., Callahan,C., Hwang,H.S., Potter,N.T., Ross,C.A. and Margolis,R.L.					

Pred. No. is the number of results predicted by chance to have a

TITLE A repeat expansion in the gene encoding junctophilin-3 is associated with Huntington disease-like 2
JOURNAL Nat. Genet. 29 (4), 377-378 (2001)

PUBMED 11694876

REFERENCE 2 (bases 1 to 125020)
AUTHORS Holmes,S.E., Ingersoll-Ashworth,R.G., Ross,C.A. and Margolis,R.L.

JOURNAL Direct Submission
Submitted (05-OCT-2001) Psychiatry, Johns Hopkins Medical Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA
Location/Qualifiers 1..125020
Source /organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/map="16q24.3"; between D16S520 and WI-12410"
/note="isolated from a patient with Huntington's Disease-Like 2 (HDL2)"
complement:(35581..35746)/rpt_type="candem"/rpt_unit="ctg/complement(<36507..>36887)
/gene="JPH3"/note="JP3"
product="Junctionophilin 3" complement(<36507..>36887)
/gene="JPH3"/product="Junctionophilin 3" complement(<36507..>36887)
/gene="JPH3"/note="component of the junctional complex between plasma membrane and endoplasmic reticulum"
/codon_start=1/product="Junctionophilin 3"
/db_xref="GI:17646245."
/translation "MSGGGFNFDDGGSYKQWQAQGRHRGTLIESKGKWWYKGENTHGFKGYGYRECAG
NGAKYEGTWTWSNGLQDGSYGTYSYSDG"
ORIGIN 29056 a 32731 c 30696 g 28283 t 4254 others

BASE COUNT

ORIGIN

Query Match 4.8%; Score 4.9.8; DB 9; Length 125020;
Best Local Similarity 9.38%; Pred. No. 0.007; Matches 41; Conservative 213; Mismatches 185; Indels 1; Gaps 1;

Db 17666 WSSYSTMGXSTSCKKYKCSWSMVKCUTSKSFRKRKYYYNGKGKRAKKKYCAGR 17607
Qy 585 GAATGTTCTCAACHTCTGCAATTTCGGTTAGCGAGTAGCTGCATTTGCATTTGTCGC 644
Db 17606 RRMSSWCKAWWMSTCWSMTCMITYYSKSTYKSTCKRGGTYWGSKTCYSGKSKS 17547
Qy 645 CTGCGAGAGACTCGTTACTGTT-CGGCGGAACGGACTCCCTCTGCTGTGCCA 703
Db 17546 MYCMRMRSSKSSWSMRSWSMCWGWAGYRRRSRSAGNRSSKSRSTOMKRACS 17487
Qy 704 ACTGCCAACCGACCTGGCATGGCTTGCGCATCTTAATTCCTCCATCARAGAAC 763
Db 17486 RKTGSGTGTGSMKKKGYSYSRGMGKKTCKCMWKYKRTSMCWWYKNSWGKFR 17427
Qy 764 ACCTCCCATATTCTCCCTGTGGCGACCTCGTGGCCCTGAAATTCCAGGGCACACGG 823
Db 17426 YKRCMCKKGKGTGYROMSSKSGYTSMRGKRGSSYSTSWMKCKYMSYKK 17367
Qy 824 TGAGGGACACCCACCCATCAAAATACGGCTCTTGGCGCAGGATCTGCACCGAGGG 883
Db 17366 RRSRMRSMSSKGWRYAGRCYSSMMWSRKRSCYSKSYKGRGMKGKRSK 17307
Qy 884 CGCAATTGAAATTCCAGGCCATGTCAGGTTGGTGGTGGCGCAGGGAAACCTGAATGGC 943
Db 17306 YWSMCKMRSSWSKCYSTSKYKSGRSTSCKSKAKSNRMAGSKCTYGSSTWSN 17247

Query Match 4.3%; Score 44.8; DB 9; Length 125020;
Best Local Similarity 10.1%; Pred. No. 0.22%; Matches 53; Conservative 207; Indels 4; Gaps 1;

Db 50579 CTRWTRBWCWSKHKSSHRMTDTYMRKTYCWDWSGY 50638
Qy 470 CAGACGAGACAATTCCGGCATCAAACAACACTTTCGGCACAGTAGTATTGGTAGGCC 529
Db 50579 CTRWTRBWCWSKHKSSHRMTDTYMRKTYCWDWSGY 50638
Qy 530 TGAACCGCTGCTATTGAGCTGGTGGTGGCGAACCGATGGATGTCCTCATCGAACCT 589
Db 50639 GGWRGMKMBDNBMBBSMMGKMSDGTDKDWCVYSSMSVDSVBYBMRWBDSWSGDBSHYMR 50638

QY	590	TGTCTCCAACTCAGCATTTGGATTCAGAGCAG --- TACTGGATTTGCATTTGTCGCC	645	* be preserved.
Db	50699	GVTGVMRKGVGBCDTHVTDGGHGSWMKSCGYSDBGCKSKMTCGMSMTKTYC	50758	* 15291: contig of 15291 bp in length * 15292 15391: gap of 100 bp * 15392 216878: contig of 201487 bp in length.
FEATURES		Location/Qualifiers		
Source		1..216878		
QY	646	TGCAAGAGACTCGTTACTGTCGCCGGAACCTCCCTGCTGTCACAA	705	"organism="Mus musculus"
Db	50759	YGTGYWSSKADESYYSBISHB/WGSTMCRCBSYASYYDARYKBGMYRVRCSWY	50818	/db_xref="taxon:10090" /chromosome="2" /clone="RP23-183012" /clone_idb="RPCI-23" 1..15291
QY	706	TGCAACAGAACCTGGCATGGCTTGGCTCATCTAATTCCTCCATCAAAGAAC	765	/note="assembly_fragment:02090 fragment_chain:1"
Db	50819	YVKYBMSVYARKSKGHVKBSSMKCSRKRKDMSTSMWRSMDRKCSRSHSYKSMGK	50878	misc_feature 15392..216878 /note="assembly_fragment:03034 fragment_chain:1"
BASE COUNT	70260	a 41471 c 41503 g 63544 t 100 others		
ORIGIN				
QY	766	CCTCCATTACTCCCCCGTGCGCGGACCTGCGGCCCTGAATTCCAGGGCACACGGTG	825	Query Match 4..3%; Score 44..4; DB 2; Length 216878; Best Local Similarity 47..5%; Pred. No. 0..31; Matches 132; Conservative 0; Mismatches 146; Indels 0; Gaps
Db	50879	WTSDVDRCYACSSBDMKHYKMRKKHDKOSHBSMGKSHMKGMVRSMYRSMCSVHDSS	50938	misc_feature
QY	826	CAGGACACACACCCGATCAAATAACGGCTCTTGGCGCACGGATCTGACCGACGGC	885	QY 177 AAGATGTTGGATTACAACAGTTCTATCTAGGGCACGCTTAACGCAGCTTGATAGTC 236 Db 137525 AAATGTCCTAAAATAGCAGAAAATTAACAGATTGACTATAATACCTCTGTATTAGTT 137578 QY 237 ATTTTATCUCGCCACCGCCATTAGTAAGCAGGCCAACGCAATTCTGATGTA 296 Db 137585 ATTGCAATTGATGACTGTCCTTATTTTAAGTATTGAAATTGATCATTAAAGGA 137574
Db	50939	MVDGSRRMKGSAARGMKYCCTYSSMRSTKSRSMKSSNFWGSCYYCYGWSCTKMRSGM	50998	QY 297 ATTTGTCATCTCATTTGGTTAGCATGTTGAACGATGTCACTTTCAATCCACATGGACAG 356 Db 137645 CTTATCATATGTCCTCCCTTATATACTAAGATATCATTAAATTTCATCCAGCCA 13770
QY	886	GCAATTGAAATCGAGCCCTGATCGAGGGAAACCTGTAATGGCC	945	QY 357 ATTCGAGAGAAATTATCCAACTCACATAAGCAGAACAGTTAACCATATAAA 416 Db 137705 ATTGTCATATAAGATTAATTCAGATTAATTAACAGAAAGCATCTGAAAGTATAAA 137776
Db	50999	CSYTGSSMKCMGRYCKSRGSMSMAGSMSMSMGRKTSMSGRCCA	51058	QY 417 CATCCGAAATTACGAGTTGTCACATTGATCTTTAGA 454 Db 137765 CATCATGAATATGGTATTAGATGCTGCTTCATAGA 137802
RESULT	3			
AL773534	216878 bp	DNA linear	HTG 17-AUG-2002	
LOCUS	Mus musculus chromosome 2 clone RP23-183012, *** SEQUENCING IN PROGRESS	2 unorderd pieces.		
DEFINITION				
AL773534	9	GI:22416034		
VERSION	HNG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.			
KEYWORDS				
SOURCE	house mouse.			
ORGANISM	Mus musculus			
Eukaryota; Mammalia; Chordata; Craniata; Vertebrata; Buteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
REFERENCE	1 (bases 1 to 216878)			
AUTHORS	Johnson,C.			
TITLE	Direct Submission			
JOURNAL	Submitted (16-AUG-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: Clonerrequest@sanger.ac.uk			
COMMENT	On Aug 21, 2002 this sequence version replaced gi:22002783.			
	----- Genome Center			
	Center: Wellcome Trust Sanger Institute			
	Center code: SC			
	Web site: http://www.sanger.ac.uk			
	Contact: humquery@sanger.ac.uk			
	----- Project Information			
	Center project name: bml83012			
	----- Summary Statistics			
	Assembly program: XGAP; version 4..5			
	Chemistry: Dye-terminator; 100% of reads			
	Consensus quality: 216691 bases at least Q40			
	Consensus quality: 216754 bases at least Q30			
	Consensus quality: 216766 bases at least Q20			
	Insert size: 216778; sum-of-contigs			
	Insert size: 211010; 5..8% error; agarose-fp			
	Quality coverage: 7..0lx in Q20 bases; sum-of-contigs Quality coverage: 7..25x in Q20 bases; agarose-fp			
*	NOTE: This is a 'working draft' sequence. It currently consists of 2 contigs. The true order of the pieces			
*	is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.			
*	as soon as it is available and the accession number will be updated with the finished sequence.			
*	as soon as it is available and the accession number will be updated with the finished sequence.			
FEATURES	Source	1..685		
		/organism="Castanea dentata"		
		/clone="AC52_33"		

misc_feature 1. .685
 BASE COUNT 253 a /note="vascular protein promoter region"
 ORIGIN

Query Match 4.2%; Score 42.8; DB 8; Length 685;
 Best Local Similarity 63.7%; Pred. No. 0.31;
 Matches 65; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

Qy 1 CTTACTATAGGGCACGGTGGTCACGCCGGCTGCTGAA 46
 Db 8 CTTACTATAGGCAGCGCTGGTCACGCCGGCTGGTCACCTCACTG 60
 Qy 61 GATGCATAATCTGTGAGAGATAACATCACAACTCACAGCT 102
 Db 68 AGATAAAAAATATTATAACTGAATAAAAATAAGCT 109

RESULT 5 AF039526 LOCUS AF039526 1423 bp DNA linear PRI 01-MAY-2001
 DEFINITION Homo sapiens MHC class I related protein 1 (MR1) gene, partial cds.
 ACCESSION AF039526
 VERSION GI:4104807
 SOURCE Homo sapiens
 ORGANISM Homo sapiens
 Mammalia; Chordata; Craniata; Vertebrata; Euteleostomi;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 Rieger,P.; Wanner,V. and Bahram,S.
 REFERENCE 1 (bases 1 to 1423)
 AUTHORS Rieger,P., Wanner,V. and Bahram,S.
 TITLE Genomics, isoforms, expression, and phylogeny of the MHC class
 I-related MR1 gene
 JOURNAL J. Immunol. 161 (8), 4066-4077 (1998)
 MEDLINE 98451457
 PUBMED 9780177
 REFERENCE 2 (bases 1 to 1423)
 AUTHORS Rieger,P., Wanner,V., Hauptmann,G. and Bahram,S.
 TITLE Direct Submission
 JOURNAL Submitted (22-DEC-1997) EG, Basel Institute for Immunology,
 Grenzacherstrasse 487, EG, 4005, Switzerland
 FEATURES Location/Qualifiers
 1..1423
 source /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="1"
 /map="1q25"
 1..>1423
 gene .1..1369
 misc_feature /gene="MR1"
 /note="Contains putative promoter and 5'UTR"
 1390..>1423
 /gene="MR1"
 /product="MHC class I related protein 1"
 /evidence="not_experimental"
 1390..>1423
 /gene="MR1"
 /codon_start=1
 /evidence="not_experimental"
 /product="MHC class I related protein 1"
 /protein_id="AAD01172.1"
 /db_xref="GI:4104808"
 /translation="MSELMLFLPL"

BASE COUNT 342 a 334 c 310 g 437 t
 ORIGIN

Query Match 4.18%; Score 42.4; DB 9; Length 1423;
 Best Local Similarity 97.7%; Pred. No. 0.47;
 Matches 43; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTTACTATAGGGCACGGTGGTCACGCCGGCTGGTATGAA 44
 Db 3 CTTACTATAGGGCACGGTGGTCACGCCGGCTGGTATGAA 46

RESULT 6 AR124194 LOCUS AR124194 Sequence 1 from patent US 6171857. DNA linear PAT 16-MAY-2001
 DEFINITION Sequence 1 from patent US 6171857.
 VERSION AR124194
 KEYWORDS Unknown
 SOURCE Unknown
 ORGANISM Unclassified
 REFERENCE 1 (bases 1 to 6078)
 AUTHORS Hendrickson,E.A.
 TITLE Leucine zipper protein, KARP-1 and methods of regulating DNA
 dependent protein kinase activity
 patent: US 6171857 A 1 09-JAN-2001;
 JOURNAL Location/Qualifiers
 FEATURES Source 1..6078
 /organism="unknown"
 BASE COUNT 1806 a 1301 c 1392 g 1543 t 36 others
 ORIGIN

Query Match 4.1%; Score 42.4; DB 6; Length 6078;
 Best Local Similarity 97.7%; Pred. No. 0.62;
 Matches 43; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTTACTATAGGGCACGGTGGTCACGCCGGCTGGTATGAA 44
 Db 9 CTTACTATAGGGCACGGTGGTCACGCCGGCTGGTATGAA 52

RESULT 7 PFMAL4P1_3/c
 WPCOMMENT

Sequence split into 4 fragments
 Fragment Name Begin End
 PFMAL4P1_0 1 110000
 PFMAL4P1_1 10001 210000
 PFMAL4P1_2 200001 310000
 PFMAL4P1_3 300001 392633
 Continuation (4 of 4) of PFMAL4P1 from base 300001 (AL034557 Plasmodium falciparum 3D

Query Match 4.1%; Score 41.8; DB 2; Length 92633;
 Best Local Similarity 65.6%; Pred. No. 1.6;
 Matches 61; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

Qy 1 CTTACTATAGGGCACGGTGGTCACGCCGGCTGGTATGAGGTGGAAACCTCAC TG 60
 Db 7315 CTTACTATAGGGCACGGTGGTCACGCCGGCTGGTATGAGGTGGAAACCTCAC TG 7256

Qy 61 GATGGATATACTGCTGAGAGATAACATCACAAAT 93
 Db 7255 TTCCCATATTAAAAAGCTCATATAATAAT 7223

RESULT 8 PFMAL4P1_2/c
 WPCOMMENT

Sequence split into 4 fragments
 Fragment Name Begin End
 PFMAL4P1_0 1 110000
 PFMAL4P1_1 10001 210000
 PFMAL4P1_2 200001 310000
 PFMAL4P1_3 300001 392633
 Continuation (3 of 4) of PFMAL4P1 from base 200001 (AL034557 Plasmodium falciparum 3D

Query Match 4.1%; Score 41.8; DB 2; Length 110000;
 Best Local Similarity 65.6%; Pred. No. 1.6;
 Matches 61; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

Qy 1 CTTACTATAGGGCACGGTGGTCACGCCGGCTGGTATGAGGTGGAAACCTCAC TG 60

Db	107315	CCTRACTATAGGGCACGCCGNGGTTGACGCCGGCCTGTGTAAATTTTTTTTT	107256
QY	61	GATGCATACTGCTGAGATAACATACAAT	93
Db	107255	TTCUCCATATTAAAAAAGCTCATATAAT	107223
RESULT	9		
AY124482/c	AY124482	4260 bp DNA linear	VRT 31-JUL-2002
LOCUS		Danio rerio myogenin gene, exons 1, 2 and 3 and complete cds.	
DEFINITION		Danio rerio myogenin gene, exons 1, 2 and 3 and complete cds.	
VERSION	AY124482	AY124482.1	GI:22035404
KEYWORDS			
SOURCE			
ORGANISM	Danio rerio		
	Fish		
	Actinopterygii; Neopterygii; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Cyprinidae; Cypriniformes; Teleostei; Ostariophysi;		
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REFERENCE			
AUTHORS	Du, S., Gao, J. and Anyangwe, V.		
TITLE	Muscle specific expression of myogenin in zebrafish embryos is controlled by multiple regulatory elements in the promoter		
UNPUBLISHED			
2	(bases 1 to 4260)		
2	(bases 1 to 4260)		
Direct Submission			
Submitted (20-JUN-2002) Center of Marine Biotechnology, University of Maryland Biotechnology Institute, 701 E. Pratt St., Baltimore, MD 21202, USA			
FEATURES			
source			
1.	.4260		
	/organism="Danio rerio"		
	/db_xref="Taxon:7955"		
PROMOTER			
mRNA			
exon			
CDS			
	/number=1		
	join(754..1309, 2288..2393, 3002..3120)		
	/codon_start=1		
	/product="myogenin"		
	/protein_id="P4AM622616_1"		
	/db_xref="GI:22035405"		
	/translation="MELFETNYYFFNDQRFYEGADNEFFQSIRNGPEQAGYDQRNSMMGLCGDGGRMLTIVGLEYKPKSPSSGLSMSPHOEQHCPCGGCLPWACKVCRKSQVMDRIKKADREKRKKVNEAFLKLMNPMLNQRLPAVEILRSIAQIYERJQALVSSLNOQEHEQNLHYTRATAAPHTCPTVSSSDQGSSESSCQSSPENASSADHCPVAYSSAHEDLNNDDSEOSNLSLTSVDTSTGTPTVAYSVDISK"		
exon			
	/number=2		
	2288..2393		
	3002..3599		
exon			
	/number=3		
BASE COUNT	1361	a 797 c	731 g 1371 t
ORIGIN			
Query Match			
Best Local Similarity	4.0%		
Matches	44	Conservative	Score 41; DB 5; Length 4260;
	0;	Mismatches 0; Indels 5; Gaps 0;	Pred. No. 1.5;
QY	2	TCTACTATAGGGCACGCCGNGGTTGACGCCGGCCTGTGTATGAGGTGG 50	PRI 31-Oct-1999
Db	4242	TCTACTATAGGGCACGCCGNGGTTGACGCCGGCCTGTGTAACTAGGCAAG 4194	sequence.
RESULT	10		
AF051769/c	AF051769	11957 bp DNA linear	PRI 31-Oct-1999
LOCUS		Homo sapiens hyaluronidase-like pseudogene 1 (HYALP1), partial	
DEFINITION			
VERSION			
ACCESSION	AF051769	GI:6164589	
	AF051769	gene	
		promoter	
REFERENCE			
AUTHORS	Vidmar, J.J., Schoerring, J.K. and Glass, A.M.D.		
TITLE			
JOURNAL			
FEATURES			
source			
2	(bases 1 to 1094)		
	/organism="Hordeum vulgare subsp. vulgare"		
	/cultivar="Klondeik-like"		
	/db_xref="Taxon:112509"		
	1..>1094		
	/genename="HVST1"		
	1..1024		
	/gene="HVST1"		
	1025..>1094		
	/organism="Hordeum vulgare subsp. vulgare"		

	source			
/gene="HVS1"	/product="high affinity sulfate transporter"			
.0225 .1089	/product="high affinity sulfate transporter"			
/gene="HVS1"	/product="high affinity sulfate transporter"			
.090 .>1094	/product="high affinity sulfate transporter"			
/codon_start=1	/product_id="AAC2664.1"			
/clone_xref="GI:3643824"	/clone="RP11-271B5"			
/translaton="MP"	/clone.lib="RPCI-11.1"			
BASE COUNT	286 a 243 g 304 t			
ORIGIN				
Query Match	4.0%			
Best Local Similarity	95.5%			
Matches	42; Conservative 0; Mismatches 2; Indels 0; Gaps 0;			
Qy	1 CTTACTATAGGCCACGCCCTGGTGACGCCGGCTGGTATGAA 44			
Db	63 CTTACTATAGGCCACGCCCTGGTGACGCCGGCTGGTATGCA 106			
RESULT 12				
AW139378/C	166703 bp DNA linear PRI 02-DEC-2000			
LOCUS	Human DNA sequence from clone RP11-271B5 on chromosome 13 Contains			
DEFINITION	a gene for a protein similar to ribosomal protein S7, the FGF9 (fibroblast growth factor 9 (glia-activating factor)) gene, ESTs, STSS, GS and CPG Islands, complete sequence.			
ACCESSION	AL139378			
VERSION	AL139378.15 GI:8247510			
KEYWORDS	HTG; CPG Island; FGF9; growth factor; ribosomal protein.			
SOURCE	human.			
ORGANISM	Homo sapiens			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 166703)			
AUTHORS	Blakey,S.			
TITLE	Direct Submission			
JOURNAL	Submitted (18-JUL-2000) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk			
COMMENT	On Jun 4, 2000 this sequence replaced gi:8246895. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.			
	This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.			
	The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:			
Em:	EMBL; Sw:	SWISSPROT; Tr:	TRINITY; Wp:	WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 13, constructed by the Sanger Centre Chromosome Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chromosome_13
CDSS	RP11-271B5 is from the library RPCI-11.1 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see http://bacpac.med.buffalo.edu/			
FEATURES	VECTOR: pBACE3.6 This sequence is the entire insert of clone RP11-271B5 The true left end of clone RP11-28A24 is at 82919 in this sequence. The true right end of clone RP11-110K8 is at 72244 in this sequence. Location/Qualifiers			
	/product="not_experimental" /product="ba271b5.1" (similar to ribosomal protein S7)" /protein_id="CAC17691.1" /db_xref="GI:11344936" /translation="MSSSAKVKPKNDPEFGSISOALLEMNSDIKAQLRITAKEJELGGRKALLIFVP:IPQLSFOKIVRLRELEKFSKGHVVFIVQRRAKPTRKRSRTNKOKRPSHTLAVHDALLELYPSVSKRTRVKLGDSRLKVKHQQNNEVHKVETESGYKKLTFQVPEERFQLQTRMT"			
	/evidence="not_experimental" /product="ba271b5.1" complement(I0958 . .11555) /gene="ba271b5.1" complement(I10953 . .11555) /gene="ba271b5.1" complement(I10953 . .11555) /note="match: cDNA: Em:MT7233 match: ESTs: Em:AA105661 Em:AT462024 match: proteins: Tr: Q02NS1 Sw: Q10101 Sw: P03362 Sw: P23821 Sw: P50894" Tr: Q29190 Tr: O13167 Sw: P50894" /codon_start=1			
	/misc_feature			
	/product="ba271b5.1" (similar to ribosomal protein S7)" /protein_id="CAC17691.1" /db_xref="GI:11344936" /translation="MSSSAKVKPKNDPEFGSISOALLEMNSDIKAQLRITAKEJELGGRKALLIFVP:IPQLSFOKIVRLRELEKFSKGHVVFIVQRRAKPTRKRSRTNKOKRPSHTLAVHDALLELYPSVSKRTRVKLGDSRLKVKHQQNNEVHKVETESGYKKLTFQVPEERFQLQTRMT"			
	/gene="ba271b5.1" /note="match: GSS: Em:AO887379" /note="match: GSS: Em:AO476700" /note="match: GSS: Em:AO593042" /note="match: GSS: Em:AO593042"			

misc_feature
 11256 . "11498 /note="match: GSS: Em:AQ264373"
 misc_feature
 11281 . "1156 /note="match: GSS: Em:AQ554409"
 repeat_region
 11672 . "11854 /note="MER5A repeat: matches 3 . 188 of consensus"
 repeat_region
 11884 . "12154 /note="L1M4 repeat: matches 2891 . 3182 of consensus"
 repeat_region
 12941 . "13266 /note="MER2 repeat: matches 1 . 345 of consensus"
 repeat_region
 13779 . "13989 /note="L1M4C repeat: matches 1511 . 1736 of consensus"
 repeat_region
 14084 . "14294 /note="L1M4C repeat: matches 1260 . 1464 of consensus"
 repeat_region
 15515 . "15827 /note="AluY repeat: matches 1 . 303 of consensus"
 repeat_region
 16476 . "16517 /note="AluS repeat: matches 3 mer tta 78% conserved"
 repeat_region
 16651 . "16946 /note="AluS repeat: matches 1 . 296 of consensus"
 repeat_region
 17442 . "17709 /note="AluSX repeat: matches 6 . 269 of consensus"
 repeat_region
 18944 . "19037 /note="MER repeat: matches 49 . 143 of consensus"
 misc_feature
 19023 . "19447 /note="match: GSS: Em:AK630744"
 misc_feature
 1936 . "19496 /note="match: GSS: Em:AK630723"
 repeat_region
 19722 . "19854 /note="MER5B repeat: matches 1 . 138 of consensus"
 repeat_region
 19863 . "20033 /note="MER5A repeat: matches 1 . 181 of consensus"
 repeat_region
 20555 . "20388 /note="L12 repeat: matches 2612 . 2750 of consensus"
 repeat_region
 23364 . "23487 /note="L12 copies 2 mer tg 100% conserved"
 repeat_region
 23300 . "23813 /note="AluSX repeat: matches 1 . 312 of consensus"
 repeat_region
 25590 . "26066 /note="L12 repeat: matches 2662 . 2739 of consensus"
 repeat_region
 26885 . "26452 /note="L12 repeat: matches 2612 . 2750 of consensus"
 repeat_region
 27300 . "27813 /note="MER5B repeat: matches 1 . 100% conserved"
 misc_feature
 27912 . "27888 /note="MER5CB repeat: matches 1 . 494 of consensus"
 repeat_region
 27912 . "27888 /note="MER5CB repeat: matches 1 . 494 of consensus"
 misc_feature
 27912 . "27888 /note="L12 repeat: matches 2597 . 2697 of consensus"
 repeat_region
 28519 . "28685 /note="AluSX repeat: matches 1 . 295 of consensus"
 repeat_region
 29089 . "29230 /note="MER5B repeat: matches 2597 . 2697 of consensus"
 misc_feature
 29406 . "29491 /note="MER5A repeat: matches 74 . 159 of consensus"
 repeat_region
 29902 . "30303 /note="L1A5 repeat: matches 5740 . 6145 of consensus"
 misc_feature
 30726 . "31152 /note="match: GSS: Em:AK559377"
 repeat_region
 31336 . "31605 /note="MER5B repeat: matches 228 . 386 of consensus"
 repeat_region
 31895 . "32195 /note="AluY repeat: matches 1 . 297 of consensus"
 repeat_region
 32453 . "32670 /note="AluY repeat: matches 1 . 303 of consensus"
 repeat_region
 32786 . "32818 /note="MER2FA repeat: matches 96 . 128 of consensus"
 repeat_region
 32813 . "33060 /note="MER2FA repeat: matches 121 . 349 of consensus"
 repeat_region
 33068 . "33143 /note="MER2FA repeat: matches 121 . 349 of consensus"

source 1. -156325 "Homo sapiens"
 /organism="Homo sapiens"
 /db_xref="Taxon:9606"
 /chromosome="20"
 /map="p12"
 /clone="RP1-13418"
 /clone_lib="RPCI-1"
 repeat_region 1145. .1261
 /note="MER5A repeat: matches 9. .136 of consensus"
 repeat_region 1278. .1446
 /note="MER3 repeat: matches 1. .171 of consensus"
 repeat_region 1686. .1846
 /note="MIR repeat: matches 29. .179 of consensus"
 repeat_region 2396. .2495
 /note="5' copies 20 mer 75% conserved"
 repeat_region 2400. .2485
 /note="43' copies 2 mer ta 77% conserved"
 repeat_region 3504. .3563
 /note="3' copies 20 mer 90% conserved"
 repeat_region 4422. .4603
 /note="HAL1 repeat: matches 519. .771 of consensus"
 repeat_region 4847. .4897
 /note="L1MEC repeat: matches 1318. .1368 of consensus"
 repeat_region 4904. .5263
 /note="L1MB8A repeat: matches 1. .603 of consensus"
 repeat_region 5275. .5520
 /note="123 copies 2 mer tt 54% conserved"
 repeat_region 6040. .6834
 /note="L1PA3 repeat: matches 5351. .6146 of consensus"
 repeat_region 7290. .7597
 /note="AluSx repeat: matches 1. .312 of consensus"
 misc_feature 8305. .8851
 /note="match: GSS: Em:B66714"
 misc_feature 8318. .8819
 /note="match: GSS: Em:AO228387"
 repeat_region 8318. .8680
 /note="L1ME3 repeat: matches 5802. .6155 of consensus"
 repeat_region 8809. .8924
 /note="MIR repeat: matches 119. .230 of consensus"
 repeat_region 9121. .9280
 /note="MIR repeat: matches 45. .206 of consensus"
 repeat_region 9296. .9446
 /note="MER5A repeat: matches 9. .180 of consensus"
 repeat_region 10122. .10887
 /note="L1MB8 repeat: matches 5379. .6171 of consensus"
 repeat_region 10898. .11043
 /note="AluB repeat: matches 159. .305 of consensus"
 repeat_region 11097. .11191
 /note="L1MB8 repeat: matches 5276. .5374 of consensus"
 repeat_region 12638. .12711
 /note="37' copies 2 mer tt 67% conserved"
 repeat_region 12921. .13046
 /note="FLAM_A repeat: matches 1. .127 of consensus"
 repeat_region 14487. .14558
 /note="L2 repeat: matches 2225. .2749 of consensus"
 repeat_region 15346. .15105
 /note="3' copies 20 mer 86% conserved"
 repeat_region 15357. .15404
 /note="24' copies 2 mer ca 95% conserved"
 repeat_region 15662. .15798
 /note="FLAM_A repeat: matches 1. .132 of consensus"
 repeat_region 15809. .15514
 /note="MIR repeat: matches 74. .183 of consensus"
 repeat_region 16324. .16400
 /note="MER5A repeat: matches 25. .106 of consensus"
 repeat_region 17365. .17492
 /note="MER5B repeat: matches 2. .133 of consensus"
 repeat_region 17481. .17553
 /note="MER5A repeat: matches 17. .100 of consensus"
 misc_feature 17529. .17978
 /note="match: GSS: Em:AO231626"
 repeat_region 17852. .18269
 /note="MSRB repeat: matches 2. .426 of consensus"
repeat_region 20049. .20099
 /note="MIR1 repeat: matches 92. .142 of consensus"
repeat_region 21187. .21455
 /note="AluSx repeat: matches 1. .275 of consensus"
 misc_feature complement(21449. .21669)
 /note="match: GSS: Em:AO033647"
 misc_feature complement(21470. .21878)
 /note="match: GSS: Em:AO52323"
repeat_region 22450. .22758
 /note="AluSC repeat: matches 1. .309 of consensus"
 misc_feature complement(24041. .24614)
 /note="match: GSS: Em:HS581M22T"
repeat_region 24106. .24383
 /note="L2 repeat: matches 2347. .2426 of consensus"
repeat_region 24384. .24681
 /note="AluSP repeat: matches 1. .299 of consensus"
 misc_feature complement(24682. .24710)
 /note="L2 repeat: matches 2426. .2454 of consensus"
repeat_region 24711. .25001
 /note="AluSX repeat: matches 4. .308 of consensus"
repeat_region 25002. .25244
 /note="L2 repeat: matches 2454. .2687 of consensus"
 misc_feature complement(26228. .26594)
 /note="THEIC repeat: matches 1. .371 of consensus"
 misc_feature complement(26328. .26777)
 /note="match: GSS: Em:AO84144"
repeat_region 26839. .27151
 /note="AluYb8 repeat: matches 1. .313 of consensus"
 misc_feature complement(27461. .27498)
 /note="19 copies 2 mer gt 84% conserved"
repeat_region 28852. .28320
 /note="LTR16A repeat: matches 159. .227 of consensus"
 misc_feature complement(30306. .30507)
 /note="match: GSS: Em:G24652"
repeat_region 30530. .30760
 /note="19 copies 2 mer at 80% conserved"
repeat_region 30812. .31214
 /note="LIME repeat: matches 5266. .5711 of consensus"
 misc_feature complement(31219. .31288)
 /note="35' copies 2 mer at 80% conserved"
repeat_region 31221. .31300
 /note="4' copies 20 mer 75% conserved"
repeat_region 31301. .31307
 /note="LIME repeat: matches 4988. .5266 of consensus"
repeat_region 31562. .31677
 /note="LIME repeat: matches 4430. .4548 of consensus"
 misc_feature complement(31971. .32384)
 /note="L1MEC repeat: matches 274. .671 of consensus"
 misc_feature complement(32087. .32597)
 /note="L1ME repeat: matches 20 mer 75% conserved"
repeat_region 32820. .32853
 /note="17' copies 2 mer tg 85% conserved"
repeat_region 33194. .33231
 /note="19' copies 2 mer tg 84% conserved"
repeat_region 33387. .34525
 /note="7' copies 77 mer 95% conserved"
 misc_feature complement(34268. .34601)
 /note="match: GSS: Em:AZ032717"
repeat_region 34715. .35116
 /note="match: GSS: Em:AQ540824"
 misc_feature complement(32180. .32607)
 /note="match: GSS: Em:AQ422766"
 misc_feature complement(32250. .33158)
 /note="match: GSS: Em:AO636331"
repeat_region 35571. .35703
 /note="MIR repeat: matches 21. .262 of consensus"
repeat_region 35743. .36053
 /note="AluY repeat: matches 1. .311 of consensus"
repeat_region 36568. .36832
 /note="AluB repeat: matches 1. .254 of consensus"
repeat_region 38453. .38751
 /note="AluSC repeat: matches 1. .300 of consensus"
repeat_region 38803. .38939

repeat_region /note="L2 repeat: matches 2577. .2710 of consensus" 39012. : 39140
 misc_feature /note="MULTI repeat: matches 339. .541 of consensus" complement(40115..40665)
 /note="match: GSS; Em:AQ424883"
 repeat_region /note="L2 repeat: matches 2605. .2750 of consensus"
 Query Match 3.9%; Score 40.6; DB 9; Length 156325;
 Best Local Similarity 55.2%; Pred. No. 4;
 Matches 79; Conservative 0; Mismatches 64; Indels 0; Gaps 0;
 Qy 335 AGTTCAATCCACATGCCAACAGAAATTCCATCACACATAAACCA 394
 Db 151285 AGTTTATTTATATACATATTTTAACTGATGAAATTAGAGAATCT 151344
 Qy 395 GCACAGAAGTTAACCATTAACATCCAAATTAAACAGACTGGCTTAGAGA 454
 Db 151345 TAATAAATGGTAACATTAATCTAAATGATCTCTTCAGTTCTAGGA 151404
 Qy 455 GGACAACCTGGGTGCAGCAG 477
 Db 151405 AATCAACCTAGGATTATATTAG 151427

RESULT 14
 LOCUS AX320057 598 bp DNA linear PAT 14 -DEC -2001
 DEFINITION Sequence 14 from Patent WO0181606.
 VERSION AX320057
 VERSION AX320057.1 GI:17901561
 KEYWORDS
 SOURCE Oryza sativa.
 ORGANISM Eukaryota; Viriplantae; Streptophytina; Embryophyta; Tracheophyta; Spermatophytina; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoidae; Oryzeae; Oryza.
 REFERENCE 1
 AUTHORS Perera,J.R.; Lu,M. and Ray,A.
 TITLE Polynucleotide sequences from rice
 JOURNAL Patent: WO 0181606-A 14 01-NOV-2001;
 FEATURES source
 1. .598
 /organism="Oryza sativa"
 /db_xref="taxon:4530"
 BASE COUNT 174 a 131 c 113 g 179 t 1 others
 ORIGIN

Query Match 3.9%; Score 40.4; DB 6; Length 598;
 Best Local Similarity 75.8%; Pred. No. 1.6;
 Matches 50; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
 Qy 4 ACTATAGGGACCGCTGGTGCGACGGCGCCGGCTGGTGTATGAAGGGAAACCTCACTGGAT 63
 Db 1 ACTATAGGGACCGCTGGTGCGACGGCGCCGGCTGGTGTATCAACTCGTCAAAATGTCGAT 60
 Qy 64 GCATAT 69
 Db 61 TAATAT 66

RESULT 15
 AF153014 1463 bp DNA linear INV 10-APR-2002
 LOCUS AF153014 Trichomonas vaginalis Tvp14 (tvp14) gene, complete cds.
 DEFINITION AF153014
 ACCESSION AF153014.1 GI:8132114
 VERSION AF153014.1
 KEYWORDS Trichomonas vaginalis.
 SOURCE Trichomonas vaginalis.
 ORGANISM Eukaryota; Parabasalidea; Trichomonadida; Trichomonas; Trichomonadinae; Trichomonas.



Result No.	Score	Query Match	Length	DB ID	Description
1	45	' 4	388	17	BH747511 SALK_0180
2	43.6	4.2	469	17	BH251364 SALK_0114
3	43.4	4.2	111	17	BH617520 SALK_0372
4	42.8	4.2	124	17	BH613783 SALK_0349
5	42.8	4.2	384	17	BH250911 SALK_0107
6	42.8	4.2	445	17	BH634101 SALK_0447

TDNA. This sequence lies within 300 bases of the 3' end of At1g57320.

Class: TDNA tagged.

FEATURES
source
 1. .388
 organism="Arabidopsis thaliana"
 strain="Columbia 0"
 /db_xref="taxon:3702"
 /clone="SALK_018023;45..85..x"
 /clone.lib="Arabidopsis thaliana TDNA insertion lines"
 /note="PCR was performed on Arabidopsis thaliana lines
each of which contains one or more TDNA insertion
elements. The resultant fragment for each line was
directly sequenced to determine the genomic sequence at
the site of insertion. Details of the protocols used can
be found at http://signal.salk.edu/tDNA_protocols.html"

the site of insertion. Details of the protocols used can
be found at http://signal.salk.edu/tDNA_protocols.html"
93 a 93 a 120 c 81 g 135 t 40 others

FEATURES	BASE COUNT	ORIGIN	Query Match	Best Local Similarity	Score	Pred.	No.	Mismatches	Indels	Gaps	0;
source	98 a	95 c	Query Match	4.4%	Score 45;	DB 17;	Length 388;				
	82 g	110 t	Best Local Similarity	78.3%	Pred. No.	0.0094;	Length 388;				
			Matches	54;	Conservative	0;	Mismatches 15;	Indels 0;	Gaps 0;		
			Qy	1	CCTACTATAGGCCACGGCTGTCACGGCCGGCTGGTATGAAGGTGGAAACCTCACTG 60						
			Db	265	CTCTCATATAGGCCACGGCTGTCACGGCCGGCTGGTATGAAGGTGGAAACAGGGAA 206						
			RESULT	2	BH251364/c	469 bp	DNA linear	GSS 28-NOV-2001			
			DEFINITION	SALK_011461	Arabidopsis thaliana TDNA insertion lines	Arabidopsis	thaliana genomic clone SALK_011461,	DNA sequence.			
			ACCESSION	BH251364							
			VERSION	BH251364..1	GI:17138342						
			KEYWORDS	GSS							
			ORGANISM	Arabidopsis thaliana	Streptophyta; Embryophyta; Tracheophyta; Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Arabidopsis; Arabidopsis.						
			REFERENCE	1 (bases 1 to 111)							
			AUTHORS	Alonso, J.M., Leisse, T.J., Barajas, P., Chen, H., Cheuk, R., Gadrinab, C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Prednis, L., Shinn, P., Zimmerman, J., and Ecker, J.R.							
			TITLE	A Sequence-Indexed Library of Insertion Mutations in the							
			JOURNAL	Unpublished (2001)							
			COMMENT	Contact: Joseph R. Ecker Salk Institute Genomic Analysis Laboratory (SIGnAL) The Salk Institute for Biological Studies 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA Tel: 858 453 4100 x1752 Fax: 858 558 6379 Email: ecker@salk.edu							
			FEATURES	Source							
			ORGANISM	Arabidopsis thaliana	Thalaina TDNA insertion lines	Arabidopsis	thaliana genome	TDNA insertion lines	1..111		
			REFERENCE	1 (bases 1 to 469)							
			AUTHORS	Alonso, J.M., Leisse, T.J., Barajas, P., Chen, H., Cheuk, R., Gadrinab, C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Prednis, L., Shinn, P., Zimmerman, J., and Ecker, J.R.							
			TITLE	A Sequence-Indexed Library of Insertion Mutations in the							
			JOURNAL	Unpublished (2001)							
			COMMENT	Contact: Joseph R. Ecker Salk Institute Genomic Analysis Laboratory (SIGnAL) The Salk Institute for Biological Studies 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA Tel: 858 453 4100 x1752 Fax: 858 558 6379							
			FEATURES	source							
			ORGANISM	Arabidopsis thaliana	TDNA insertion lines	Arabidopsis	thaliana genome	TDNA insertion lines	1..111		
			REFERENCE	1 (bases 1 to 111)							
			AUTHORS	Alonso, J.M., Leisse, T.J., Barajas, P., Chen, H., Cheuk, R., Gadrinab, C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Prednis, L., Shinn, P., Zimmerman, J., and Ecker, J.R.							
			TITLE	A Sequence-Indexed Library of Insertion Mutations in the							
			JOURNAL	Unpublished (2001)							
			COMMENT	Contact: Joseph R. Ecker Salk Institute Genomic Analysis Laboratory (SIGnAL) The Salk Institute for Biological Studies 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA Tel: 858 453 4100 x1752 Fax: 858 558 6379							
			FEATURES	source							
			ORGANISM	Arabidopsis thaliana	TDNA insertion lines	Arabidopsis	thaliana genome	TDNA insertion lines	1..111		
			REFERENCE	1 (bases 1 to 111)							
			AUTHORS	Alonso, J.M., Leisse, T.J., Barajas, P., Chen, H., Cheuk, R., Gadrinab, C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Prednis, L., Shinn, P., Zimmerman, J., and Ecker, J.R.							
			TITLE	A Sequence-Indexed Library of Insertion Mutations in the							
			JOURNAL	Unpublished (2001)							
			COMMENT	Contact: Joseph R. Ecker Salk Institute Genomic Analysis Laboratory (SIGnAL) The Salk Institute for Biological Studies 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA Tel: 858 453 4100 x1752 Fax: 858 558 6379							
			FEATURES	source							
			ORGANISM	Arabidopsis thaliana	TDNA insertion lines	Arabidopsis	thaliana genome	TDNA insertion lines	1..111		
			REFERENCE	1 (bases 1 to 111)							
			AUTHORS	Alonso, J.M., Leisse, T.J., Barajas, P., Chen, H., Cheuk, R., Gadrinab, C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Prednis, L., Shinn, P., Zimmerman, J., and Ecker, J.R.							
			TITLE	A Sequence-Indexed Library of Insertion Mutations in the							
			JOURNAL	Unpublished (2001)							
			COMMENT	Contact: Joseph R. Ecker Salk Institute Genomic Analysis Laboratory (SIGnAL) The Salk Institute for Biological Studies 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA Tel: 858 453 4100 x1752 Fax: 858 558 6379							
			FEATURES	source							
			ORGANISM	Arabidopsis thaliana	TDNA insertion lines	Arabidopsis	thaliana genome	TDNA insertion lines	1..111		
			REFERENCE	1 (bases 1 to 111)							
			AUTHORS	Alonso, J.M., Leisse, T.J., Barajas, P., Chen, H., Cheuk, R., Gadrinab, C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Prednis, L., Shinn, P., Zimmerman, J., and Ecker, J.R.							
			TITLE	A Sequence-Indexed Library of Insertion Mutations in the							
			JOURNAL	Unpublished (2001)							
			COMMENT	Contact: Joseph R. Ecker Salk Institute Genomic Analysis Laboratory (SIGnAL) The Salk Institute for Biological Studies 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA Tel: 858 453 4100 x1752 Fax: 858 558 6379							
			FEATURES	source							
			ORGANISM	Arabidopsis thaliana	TDNA insertion lines	Arabidopsis	thaliana genome	TDNA insertion lines	1..111		
			REFERENCE	1 (bases 1 to 111)							
			AUTHORS	Alonso, J.M., Leisse, T.J., Barajas, P., Chen, H., Cheuk, R., Gadrinab, C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Prednis, L., Shinn, P., Zimmerman, J., and Ecker, J.R.							
			TITLE	A Sequence-Indexed Library of Insertion Mutations in the							
			JOURNAL	Unpublished (2001)							
			COMMENT	Contact: Joseph R. Ecker Salk Institute Genomic Analysis Laboratory (SIGnAL) The Salk Institute for Biological Studies 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA Tel: 858 453 4100 x1752 Fax: 858 558 6379							
			FEATURES	source							
			ORGANISM	Arabidopsis thaliana	TDNA insertion lines	Arabidopsis	thaliana genome	TDNA insertion lines	1..111		
			REFERENCE	1 (bases 1 to 111)							
			AUTHORS	Alonso, J.M., Leisse, T.J., Barajas, P., Chen, H., Cheuk, R., Gadrinab, C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Prednis, L., Shinn, P., Zimmerman, J., and Ecker, J.R.							
			TITLE	A Sequence-Indexed Library of Insertion Mutations in the							
			JOURNAL	Unpublished (2001)							
			COMMENT	Contact: Joseph R. Ecker Salk Institute Genomic Analysis Laboratory (SIGnAL) The Salk Institute for Biological Studies 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA Tel: 858 453 4100 x1752 Fax: 858 558 6379							
			FEATURES	source							
			ORGANISM	Arabidopsis thaliana	TDNA insertion lines	Arabidopsis	thaliana genome	TDNA insertion lines	1..111		
			REFERENCE	1 (bases 1 to 111)							
			AUTHORS	Alonso, J.M., Leisse, T.J., Barajas, P., Chen, H., Cheuk, R., Gadrinab, C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Prednis, L., Shinn, P., Zimmerman, J., and Ecker, J.R.							
			TITLE	A Sequence-Indexed Library of Insertion Mutations in the							
			JOURNAL	Unpublished (2001)							
			COMMENT	Contact: Joseph R. Ecker Salk Institute Genomic Analysis Laboratory (SIGnAL) The Salk Institute for Biological Studies 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA Tel: 858 453 4100 x1752 Fax: 858 558 6379							
			FEATURES	source							
			ORGANISM	Arabidopsis thaliana	TDNA insertion lines	Arabidopsis	thaliana genome	TDNA insertion lines	1..111		
			REFERENCE	1 (bases 1 to 111)							
			AUTHORS	Alonso, J.M., Leisse, T.J., Barajas, P., Chen, H., Cheuk, R., Gadrinab, C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Prednis, L., Shinn, P., Zimmerman, J., and Ecker, J.R.							
			TITLE	A Sequence-Indexed Library of Insertion Mutations in the							
			JOURNAL	Unpublished (2001)							
			COMMENT	Contact: Joseph R. Ecker Salk Institute Genomic Analysis Laboratory (SIGnAL) The Salk Institute for Biological Studies 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA Tel: 858 453 4100 x1752 Fax: 858 558 6379							
			FEATURES	source							
			ORGANISM	Arabidopsis thaliana	TDNA insertion lines	Arabidopsis	thaliana genome	TDNA insertion lines	1..111		
			REFERENCE	1 (bases 1 to 111)							
			AUTHORS	Alonso, J.M., Leisse, T.J., Barajas, P., Chen, H., Cheuk, R., Gadrinab, C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Prednis, L., Shinn, P., Zimmerman, J., and Ecker, J.R.							
			TITLE	A Sequence-Indexed Library of Insertion Mutations in the							
			JOURNAL	Unpublished (2001)							
			COMMENT	Contact: Joseph R. Ecker Salk Institute Genomic Analysis Laboratory (SIGnAL) The Salk Institute for Biological Studies 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA Tel: 858 453 4100 x1752 Fax: 858 558 6379							
			FEATURES	source							
			ORGANISM	Arabidopsis thaliana	TDNA insertion lines	Arabidopsis	thaliana genome	TDNA insertion lines	1..111		
			REFERENCE	1 (bases 1 to 111)							
			AUTHORS	Alonso, J.M., Leisse, T.J., Barajas, P., Chen, H., Cheuk, R., Gadrinab, C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Prednis, L., Shinn, P., Zimmerman, J., and Ecker, J.R.							
			TITLE	A Sequence-Indexed Library of Insertion Mutations in the							
			JOURNAL	Unpublished (2001)							
			COMMENT	Contact: Joseph R. Ecker Salk Institute Genomic Analysis Laboratory (SIGnAL) The Salk Institute for Biological Studies 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA Tel: 858 453 4100 x1752 Fax: 858 558 6379							
			FEATURES	source							
			ORGANISM	Arabidopsis thaliana	TDNA insertion lines	Arabidopsis	thaliana genome	TDNA insertion lines	1..111		
			REFERENCE	1 (bases 1 to 111)							
			AUTHORS	Alonso, J.M., Leisse, T.J., Barajas, P., Chen, H., Cheuk, R., Gadrinab, C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Prednis, L., Shinn, P., Zimmerman, J., and Ecker, J.R.							
			TITLE	A Sequence-Indexed Library of Insertion Mutations in the							
			JOURNAL	Unpublished (2001)							
			COMMENT	Contact: Joseph R. Ecker Salk Institute Genomic Analysis Laboratory (SIGnAL) The Salk Institute for Biological Studies 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA Tel: 858 453 4100 x1752 Fax: 858 558 6379							
			FEATURES	source							
			ORGANISM	Arabidopsis thaliana	TDNA insertion lines	Arabidopsis	thaliana genome	TDNA insertion lines	1..111		
			REFERENCE	1 (bases 1 to 111)							
			AUTHORS	Alonso, J.M., Leisse, T.J., Barajas, P., Chen, H., Cheuk, R., Gadrinab, C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Prednis, L., Shinn, P., Zimmerman, J., and Ecker, J.R.							
			TITLE	A Sequence-Indexed Library of Insertion Mutations in the							
			JOURNAL	Unpublished (2001)							
			COMMENT	Contact: Joseph R. Ecker Salk Institute Genomic Analysis Laboratory (SIGnAL) The Salk Institute for Biological Studies 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA Tel: 858 453 4100 x1752 Fax: 858 558 6379							
			FEATURES	source							
			ORGANISM	Arabidopsis thaliana	TDNA insertion lines	Arabidopsis	thaliana genome	TDNA insertion lines	1..111		
			REFERENCE	1 (bases 1 to 111)							
			AUTHORS	Alonso, J.M., Leisse, T.J., Barajas, P., Chen, H., Cheuk, R., Gadrinab, C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Prednis, L., Shinn, P., Zimmerman, J., and Ecker, J.R.							
			TITLE	A Sequence-Indexed Library of Insertion Mutations in the							
			JOURNAL	Unpublished (2001)							
			COMMENT	Contact: Joseph R. Ecker Salk Institute Genomic Analysis Laboratory (SIGnAL) The Salk Institute for Biological Studies 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA Tel: 858 453 4100 x1752 Fax: 858 558 6379							
			FEATURES	source							
			ORGANISM	Arabidopsis thaliana	TDNA insertion lines	Arabidopsis	thaliana genome	TDNA insertion lines	1..111		
			REFERENCE	1 (bases 1 to 111)							

TDNA. This sequence lies within an annotated exon of At1g16780.
 Class: TDNA tagged.
 Location/Qualifiers
 1. .445

FEATURES source /organism="Arabidopsis thaliana"
 /strain="Columbia 0"
 /db_xref="taxon:702"
 /clone="SALK_046918"
 /clone_lib="Arabidopsis thaliana TDNA insertion lines"
 /note="PCR was performed on Arabidopsis thaliana TDNA insertion lines, each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at http://signal.salk.edu/tDNA_protocols.html"

BASE COUNT 114 a 129 c 86 g 115 t 1 others
 ORIGIN

Query Match 4.2%; Score 42.8; DB 17; Length 445;
 Best Local Similarity 75.7%; Pred. No. 0.049;
 Matches 53; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 1 CCTACTATAGGGACGGCTGGTCGACGGCCGGCTCGTATGAAGGTGGAACCTCACTG 60
 Db 183 CCTACTATAGGGACGGCTGGTCGACGGCCGGCTGCATAAACCTAACCTCACTC 124

QY 61 GATCCATA 70
 Db 123 ACTCTATA 114

RESULT 7 BH748887 LOCUS SALK_046918.54.75.x DNA linear GSS 27-FEB-2002 DEFINITION Arabidopsis thaliana genomic clone SALK_046918.54.75.x, DNA sequence.
 ACCESSION BH748887
 KEYWORDS GSS
 SOURCE thale cress.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicots; Rosidae; eurosids II; Brassicales; Brassiceae; Arabidopsis.
 REFERENCE AUTHORS Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P., Zimmerman,J., and Ecker,J.R.
 TITLE A Sequence-Indexed Library of Insertion Mutations in the JOURNAL Unpublished (2001)
 COMMENT Contact: Joseph R. Ecker
 Salk Institute Genomic Analysis Laboratory (SIGnAL)
 The Salk Institute for Biological Studies
 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
 Tel: 858 453 4100 x1752
 Fax: 858 558 6379
 Email: ecker@salk.edu
 This is single pass sequence recovered from the left border of TDNA.
 Class: TDNA tagged.

FEATURES source
 1. .440

FEATURES source /organism="Arabidopsis thaliana"
 /strain="Columbia 0"
 /db_xref="taxon:3702"
 /clone="SALK_046918.54.75.x"
 /note="PCR was performed on Arabidopsis thaliana TDNA insertion lines, each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at http://signal.salk.edu/tDNA_protocols.html"

BASE COUNT 112 a 106 c 95 g 111 t 8 others
 ORIGIN

Query Match 4.1%; Score 42.2; DB 17; Length 432;
 Best Local Similarity 74.6%; Pred. No. 0.075;
 Matches 53; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 4 ACTATAGGCCACGGCTGGTCGAGGCCGGCTGGTATGAGGTGGCAACCTCACTGAT 63
 Db 259 ACTATAGGCCACGGCTGGTCGAGGCCGGCTGGTATGAGGTGGCAACCTCACTGATCATGGTGTGACACTA 200

the site of insertion. Details of the protocols used can be found at http://signal.salk.edu/tDNA_protocols.html"

BASE COUNT 129 a 96 c 111 g 104 t
 ORIGIN

Query Match 4.1%; Score 42.6; DB 17; Length 440;
 Best Local Similarity 67.4%; Pred. No. 0.057;
 Matches 60; Conservative 29; Mismatches 29; Indels 0; Gaps 0;

QY 1 CCTACTATAGGGACGGCTGGTCGACGGCCGGCTGGPATGAAGGTGGAACCTCACTG 60
 Db 329 CCTACTATAGGGACGGCTGGTCGACGGCCGGCTGCAGCTTGCGGACGATCTCTG 270

QY 61 GATGATAATACTGTGAGATAACATCA 89
 Db 269 GATACGTACCCCTGCGATAATTACA 241

RESULT 8 BH610787/c LOCUS BH610787 DEFINITION SALK_018043 Arabidopsis thaliana TDNA insertion lines Arabidopsis thaliana genomic clone SALK_018043, DNA sequence.
 ACCESSION BH610787
 VERSION BH610787.1 GI:18057909
 SOURCE GSS.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicots; Rosidae; eurosids II; Brassicales; Brassiceae; Arabidopsis.
 REFERENCE AUTHORS Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P., Zimmerman,J., and Ecker,J.R.
 TITLE A Sequence-Indexed Library of Insertion Mutations in the JOURNAL Unpublished (2001)
 COMMENT Contact: Joseph R. Ecker
 Salk Institute Genomic Analysis Laboratory (SIGnAL)
 The Salk Institute for Biological Studies
 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
 Tel: 858 453 4100 x1752
 Fax: 858 558 6379
 Email: ecker@salk.edu
 This is single pass sequence recovered from the left border of TDNA. This sequence lies within 300 bases of the 5' end of At2g24510.
 Class: TDNA tagged.

FEATURES source
 1. .432
 /organism="Arabidopsis thaliana"
 /strain="Columbia 0"
 /db_xref="taxon:3702"
 /clone="SALK_018043"
 /note="PCR was performed on Arabidopsis thaliana TDNA insertion lines, each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at http://signal.salk.edu/tDNA_protocols.html"

BASE COUNT 112 a 106 c 95 g 111 t 8 others
 ORIGIN

Query Match 4.1%; Score 42.2; DB 17; Length 432;
 Best Local Similarity 74.6%; Pred. No. 0.075;
 Matches 53; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 4 ACTATAGGCCACGGCTGGTCGAGGCCGGCTGGTATGAGGTGGCAACCTCACTGAT 63
 Db 259 ACTATAGGCCACGGCTGGTCGAGGCCGGCTGGTATGAGGTGGCAACCTCACTGATCATGGTGTGACACTA 200

QY	64	GCATATACTGC	74	SOURCE	Drosophila melanogaster.
Db	199	GAATAAACCTC	189	ORGANISM	Drosophila melanogaster Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydriidae; Drosophilidae; Drosophila.
RESULT	9			REFERENCE	seqref@genoscope.cns.fr;
BH214178/C	BH214178	500 bp	DNA linear	JOURNAL	BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
DEFINITION	SALK_010362	Arabidopsis thaliana TDNA insertion lines	AraIDopsis	COMMENT	Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC Library was prepared by Kazutoyo Osoegawa and Aaron Mammone in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm .
ACCESSION	BH214178			FEATURES	Location/Qualifiers
VERSION	BH214178.1			source	1..1101 /organism="Drosophila melanogaster" /db_xref="taxon:7227" /clone="PACR30M12" /clone_lib="RPCI-98" /note="end : T7" /note="end : T7"
KEYWORDS	GSS			BASE COUNT	337 a 227 c 208 g 266 t 63 others
ORGANISM	Arabidopsis thaliana			ORIGIN	
Eukaryota; Viridiplanteae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliopsida; eudicots; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.	Query Match	4.1%	Score 42.2;	DB 17;	Length 1101;
AUTHORS	Alonso, J.M., Leise, T.J., Barajas, P., Chen, H., Cheuk, R., Gadrinab, C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Prednis, L., Shinn, P., Zimmerman, J., and Ecker, J.R.	Best Local Similarity 44.8%;	Pred. No. 0..11;	Mismatches 95;	Indels 0;
TITLE	A Sequence-Indexed Library of Insertion Mutations in the Arabidopsis Genome	Matches 90;	Conservative 16;	Gaps 0;	
JOURNAL	Unpublished (2001)				
COMMENT	Contact: Joseph R. Ecker Salk Institute Genomic Analysis Laboratory (SIGNAL) The Salk Institute for Biological Studies 1010 N. Torrey Pines Road, La Jolla, CA 92037, USA Tel: 858 453 4100 x1752 Fax: 858 558 6379 Email: ecker@salk.edu				
This is single pass sequence recovered from the left border of TDNA. This sequence lies within an exon of At5g22340. Class: TDNA tagged.	Query Match	4.1%	Score 42.2;	DB 17;	Length 1101;
FEATURES	Location/Qualifiers			Best Local Similarity 44.8%;	
source	1..500			Pred. No. 0..11;	
/organism="Arabidopsis thaliana" /strain="Columbia 0" /strain="Colombia 0" /db_xref="taxon:3702" /clone="SALK_010362" /clone.lib="Arabidopsis thaliana TDNA insertion lines" /note="PCR was performed on Arabidopsis thaliana lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at http://signal.salk.edu/tDNA_protocols.html "	Matches 90;		Mismatches 95;		
BASE COUNT	54 a	260 c	41 g	Indels 0;	
ORIGIN				Gaps 0;	
Query Match	4.1%	Score 42.2;	DB 17;	Length 500;	
Best Local Similarity 79.4%;	Pred. No. 0..08;				
Matches 50;	Conservative 0;	Mismatches 13;	Indels 0;	Gaps 0;	
Query	1	CCTACTATGGCAAGCGTGTGACGCCGGCTGATGAAGGTGGAACTCTACTG	60	Indels 0;	
Db	141	CTCACTATGGCAAGCGTGTGACGCCGGCTGATGAAGGTGGAACTCTACTG	82	Gaps 0;	
DEFINITION	Drosophila melanogaster genome survey sequence T7 end of BAC: BACR30M12 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.	Query Match	4.1%	Score 42.2;	DB 17;
VERSION	AL068402	QY 61 GAT 63	Best Local Similarity 79.4%;	Pred. No. 0..08;	Length 500;
KEYWORDS	Db 81 GAT 79	Db 1036 ATTAATTGGCAAAGAAAAGMAAA 1056	Matches 50;	Conservative 0;	
RESULT	10	CNS001LGU	1101 bp	DNA linear	GSS 03-JUN-1999
LOCUS	CNS001LGU	Drosophila melanogaster genome survey sequence T7 end of BAC: BACR30M12 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.	DEFINITION	BH168812	129 bp DNA linear GSS 03-OCT-2001
DEFINITION	AL068402	1 (bases 1 to 129)	LOCUS	BH168812	Arabidopsis thaliana TDNA insertion lines Arabidopsis.
ACCESSION	AL068402.1	Alonso, J.M., Leisse, T.J., Barajas, P., Chen, H., Cheuk, R., Gadrinab, C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Prednis, L., Shinn, P.	DEFINITION	BH168812	Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
VERSION	QSS.		ACCESSION	BH168812.1	Accession G1:15904187
KEYWORDS			VERSION	BH168812.1	Source GSS.
Organism	Arabidopsis thaliana	Organism Arabidopsis thaliana	Organism Arabidopsis thaliana		
Eukaryota; Viridiplanteae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eu dicots; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.	Eukaryota; Viridiplanteae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eu dicots; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.	Eukaryota; Viridiplanteae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eu dicots; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.	Eukaryota; Viridiplanteae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eu dicots; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.		

TITLE	Zimmerman, J. and Ecker,J.R. Unpublished COMMENT	A Sequence-Indexed Library of Insertion Mutations in the Arabidopsis Genome	FEATURES source	FEATURES source
JOURNAL	Salk Institute for Biological Studies 1001 N. Torrey Pines Road, La Jolla, CA 92037, USA Tel: 858 453 4100 x1752 Fax: 858 558 6379 Email: ecker@salk.edu	This is single pass sequence recovered from the left border of TDNA. Class: TDNA tagged.	Location/Qualifiers 1. organism="Arabidopsis thaliana" /strain="Columbia 0" /db_xref="taxon:3702" /clone="SALK_011803" /clone.lib="Arabidopsis thaliana TDNA insertion lines" /note="PCR was performed on Arabidopsis thaliana lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at http://signal.salk.edu/tDNA_protocols.html "	Location/Qualifiers 1. organism="Arabidopsis thaliana" /strain="Columbia 0" /db_xref="taxon:3702" /clone="SALK_011803" /clone.lib="Arabidopsis thaliana TDNA insertion lines" /note="PCR was performed on Arabidopsis thaliana lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at http://signal.salk.edu/tDNA_protocols.html "
FEATURES source	1..129	Location/Qualifiers	Query Match 4.1%; Score 42; DB 17; Length 220; Best Local Similarity 64.3%; Pred. No. 0 0.054; GSS 24-OCT-2001 Matches 63; Conservative 0; Mismatches 35; Indels 0; Gaps 0;	Query Match 4.1%; Score 42; DB 17; Length 220; Best Local Similarity 73.0%; Pred. No. 0 0.067; GSS 24-OCT-2001 Matches 54; Conservative 0; Mismatches 20; Indels 0; Gaps 0;
DEFINITION	/clone="SALK_000216"			
ACCESSION	1 CTTACTATAGGGCACGCGTGGCTGGCTGACGCCCGGGCTGTATGAAGGTTGGAACCTCTACTG 60		1 CTTACTATAGGGCACGCGTGGCTGGCTGGCTGACGCCCGGGCTGTATGAAGGTTGGAACCTCTACTG 60	
VERSION	Db 123 CTCACATAGGCACGGTTGTCAGGGCGGGCAATTTCGGTTGAAATGAGAG 64		Db 214 CTCACATAGGGACGCTGGCTGACGCCCGGGCTGCAATTCAAGCTAAACACAATT 155	
SOURCE	61 GATGCCATATACTGC 74		61 GATGCCATATACTGC 74	
ORGANISM	Db 25 a 38 c 28 g 38 t		Db 154 TTTTTAATTAGTGC 141	
RESULT 13	BH212774/c		BH212774	452 bp DNA linear GSS 24-OCT-2001
LOCUS			SALK_008030	Arabidopsis thaliana TDNA insertion lines Arabidopsis
DEFINITION			thaliana genomic clone SALK_008030,	DNA sequence.
ACCESSION	1 CTTACTATAGGGCACGCGTGGCTGGCTGACGCCCGGGCTGTATGAAGGTTGGAACCTCTACTG 60		BH212774	
VERSION	Db 123 CTCACATAGGCACGGTTGTCAGGGCGGGCAATTTCGGTTGAAATGAGAG 64		Db 214 CTCACATAGGGACGCTGGCTGACGCCCGGGCTGCAATTCAAGCTAAACACAATT 155	
SOURCE	61 GATGCCATATACTGC 74		61 GATGCCATATACTGC 74	
ORGANISM	Db 63 AAGAAACAGAACCAAGGAGAACTGTTAACAA 26		Db 154 TTTTTAATTAGTGC 141	
RESULT 12	BH251572/c		BH251572	220 bp DNA linear GSS 28-NOV-2001
DEFINITION	SALK_011803 Arabidopsis thaliana TDNA insertion lines Arabidopsis thaliana genomic clone SALK_011803, DNA sequence.		SALK_011803	Arabidopsis thaliana TDNA insertion lines
ACCESSION	BH251572		BH251572.1	
VERSION	GR:17138550			
SOURCE	GSS.			
ORGANISM	Arabidopsis thaliana		Arabidopsis thaliana	
DEFINITION	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.		Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.	
ACCESSION	1 (bases 1 to 220)		1 (bases 1 to 220)	
VERSION	Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab,C., Jeste,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P.		Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab,C., Jeste,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P.	
SOURCE	Zimmerman,J. and Ecker,J.R.		Zimmerman,J. and Ecker,J.R.	
DEFINITION	A Sequence-Indexed Library of Insertion Mutations in the Arabidopsis Genome		A Sequence-Indexed Library of Insertion Mutations in the Arabidopsis Genome	
VERSION	Unpublished (2001)		Unpublished (2001)	
SOURCE	Contact: Joseph R. Ecker Salk Institute Genomic Analysis Laboratory (SIGNAL) The Salk Institute for Biological Studies 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA Tel: 858 453 4100 x1752 Fax: 858 558 6379 Email: ecker@salk.edu		Contact: Joseph R. Ecker Salk Institute Genomic Analysis Laboratory (SIGNAL) The Salk Institute for Biological Studies 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA Tel: 858 453 4100 x1752 Fax: 858 558 6379 Email: ecker@salk.edu	
COMMENT	This is single pass sequence recovered from the left border of TDNA. Class: TDNA tagged.		This is single pass sequence recovered from the left border of TDNA. Class: TDNA tagged.	
JOURNAL	114 a 113 c 111 g 93 t 21 others		114 a 113 c 111 g 93 t 21 others	

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

OM nucleic - nucleic search, using sw model

Run on: December 27, 2002, 23:39:28 ; Search time 2660 Seconds
 (without alignments)
 11269.132 Million cell updates/search

Title: US-09-945-376-3
 Perfect score: 1030
 Sequence: 1 cttaataggcacgtg.....tgtataatacaaggcttg 1030

Scoring table: IDENTITY_NUC
 Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

Database : GenEmbl:
 1: gb_ba: *
 2: gb_htg: *
 3: gb_in: *
 4: gb_om: *
 5: gb_ov: *
 6: gb_pat: *
 7: gb_ph: *
 8: gb_pi: *
 9: gb_pr: *
 10: gb_ro: *
 11: gb_sts: *
 12: gb_sy: *
 13: gb_un: *
 14: gb_vt: *
 15: em_ba: *
 16: em_fun: *
 17: em_hum: *
 18: em_in: *
 19: em_mu: *
 20: em_om: *
 21: em_or: *
 22: em_ov: *
 23: em_pat: *
 24: em_ph: *
 25: em_pl: *
 26: em_rx: *
 27: em_sts: *
 28: em_un: *
 29: em_vt: *
 30: em_htg_hum: *
 31: em_htg_inv: *
 32: em_htg_other: *
 33: em_htg_mus: *
 34: em_htg_pln: *
 35: em_htg_rod: *
 36: em_htg_nam: *
 37: em_htg_vrt: *
 38: em_sy: *
 39: em_htgo_hum: *
 40: em_htgo_mus: *
 41: em_htgo_other: *

Pred. No. is the number of results predicted by chance to have a

Pred. No. is the number of results predicted by chance to have a

TITLE	A repeat expansion in the gene encoding junctophilin-3 is associated with Huntington disease-like 2	QY	944 CGGCATCAGTAGCCACGGCA 963
JOURNAL	Nat. Genet. 29 (4), 377-378 (2001)	Db	17246 NRNNRNGKTGNYMTRSR 17227
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS	Holmes,S.E., Ing尔斯oll-Ashworth,R.G., Ross,C.A. and Margolis,R.L.	RESULT	2
TITLE	Direct Submission	LOCUS	AF429315 125020 bp DNA linear
JOURNAL	Submitted (05-OCT-2001) Psychiatry, Johns Hopkins Medical Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA	DEFINITION	Vertebrata; Craniata; Chordata; Mammalia; Eutheria; Homo sapiens junctophilin 3 (JPH3) gene, partial cds.
CHARACTERS	Location/Qualifiers	ACCESSION	AF429315
SOURCE	1_125020	VERSION	AF429315.1 GI:17666244
	/organism="Homo sapiens"	KEYWORDS	
	/db_xref="taxon:9606"	SOURCE	Homo sapiens
	/chromosome="16"	ORGANISM	Homo sapiens
	/map="16q24.3; between D16S520 and WI-12410"	DEFINITION	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi; Hominidae; Homo.
repeat_region	/note="isolated from a patient with Huntington's Disease-Like 2 (HDL2)"	DEFINITION	1 (bases 1 to 125020)
	complement(35581..35746)	AUTHORS	Holmes,S.E., O'Hearn,E., Rosenblatt,A., Callahan,C., Hwang,H.S., Ingersoll-Ashworth,R.G., Fleisher,A., Stevanin,G., Brice,A., Potter,N.T., Ross,C.A. and Margolis,R.L.
gene	/rpt_unit=tctg	TITLE	A repeat expansion in the gene encoding junctophilin-3 associated with Huntington disease-like 2
	complement(<36507..>36887)	JOURNAL	Nat. Genet. 29 (4), 377-378 (2001)
mRNA	/note="JPH3"	JOURNAL	
	/product="junctophilin 3"	FEATURES	
CDS	/complement(<36507..>36887)	source	
	/gene="JPH3"	repeat_region	
	/note="component of the junctional complex between plasma membrane and endoplasmic reticulum"	gene	
	/codon_start=1	gene	
	/product="junctophilin 3"	gene	
	/protein_id="PAA40941.1"	gene	
	/db_xref="GI:17666245"	gene	
	/translation="MSSSGREFDGGSYCGGWEDGKAHGHHGVCTGPKGOCGYTGSWS	repeat_region	
	HGAEVILGVTTWISGNNTQCTWAGKQRHIGLESLKGKRWVYKGSEWTHFKGRYGVRECAG	note	"note" isolated from a patient with Huntington's Disease-Like 2 (HDL2)"
	NGAKYVLSNGLQDGIGETW2"	note	
SE COUNT	29056 a 32731 c 30696 g 28283 t 4254 others	note	
IGIN		note	
Query Match	4.8%; Score 49.8; DB 9; Length 125020;	note	
Best Local Similarity	9.3%; Pred. No. 0.007;	note	
Matches	23; Mismatches 41; Conservative 23;	note	
	Indels 1; Gaps 1;	mRNA	
525	AGGCCGTGAACCCTGTCATTGCGTATTGGAGTTGGGTGGTGCGACCCGANGGATGCCATCC 584	CDS	
17666	W\$STSTMG\$YSS\$TSCKYKCSNSMYYCKTSSYRRRSYYWGGSKTKYAGGR 17607		
585	GAACTGTGTCACCTCGCATGGTCAAGGAGTACTGCAATTGGCCATTGTCGC 644		
17606	RMSYWKSCAKWNMSYCCWMSYCKTSSYKSCITYKSSICYKRGGYNGWSKTKYAGGR 17547		
645	CTGACAGAAGACCTCGTACTGTT-CGCCGGAACCGACCTCCCTCTGCTGTCGCCA 703		
17546	MYCMMRSSKSSWMSMARSSSCIGMGAGYRFSKRSAGWAGWRSSKGKTSGMKRACS 17487		
704	ACTGCAACCAGACTGCGATGGCTGGCTCACTTAATTCTCTAICAAGAATC 763		
17486	KKTSYGTGTRSKRKKKKYSKYRGKMKKKKKYKTYKRTSNCWVYMSGYKR 17427		
764	ACCTCCCCATTACTCCGCCCTGTCGGCCGACCTCTGCGCTCAATTCCACCGAACGG 823	BASE COUNT	29056 a 32731 c 30696 g 28283 t 4254 others
17426	YRKRCMKKRKSSKCVSTKSKYRGMSKSKSYKSNRGMRSSESYSPSCWRSW31SWMKMKSYYKK 17367	ORIGIN	
824	TGCAAGCAGACACCCACGCCATCAAATAAGCCGCTCTGGGCCAGGATCTGCCAGACGG 883	Query Match	4.3%; Score 44.8; DB 9; Length 125020;
17366	RRSNRGMSSKKGIRGYAGRGYAGRCYSSMSTRRKSKCYSKYSKGRKMKWGGMKRGSK 17307	Best Local Similarity	10.1%; Pred. No. 0.22;
		Matches	53; Conservative 260; Mismatches 207; Indels 4; Gaps 1;
884	GCGCAATTGAATTGAGGTGATCGAGGTGGTGGATGCCAGGGAAACCTTGAATGGC 943	QY	470 CAGACGAGACAAATTGGGCAATCAACAACCTTGGGTGGCCACCGATGTCACCCGAACCT 529
17306	YSSSMKMKKKRKSSKCVSTKSKYRGMSKSKSYKSNRGMRSSESYSPSCWRSW31SWMKMKSYYKK 17247	Db	50579 CTGRNTGKBWSSKHSCHRTMDYRMRKTYCCCTAYMMBHYCYDBCMGVNGSY 50638
		QY	530 TGAACCGCTGGCTATTGCACTTGGGGTTOGGTGGCCACCGATGTCACCCGAACCT 5699
		Db	50639 GWRRNMMKWDWBMBSMGKSMGTDKDYWCMVMSYDVYBWRWBDTSWSRSHYR 50699

Qy	590	TGTCCTCAACTCTGATATTGGTTCAGACAG ---TAATGCATTGCGCATTTGTCGCC	645
Db	50699	GVHTGVMRWYGYGBCDTHVYTDDGHGHSWMMKSCGYSDBEGCKAKBGSRMKTGSMTKYIC	50758
Qy	646	TGCACGAGAGACTCGTTACTGTTCGGCCGGAAACGGACCTCCCTCTGCTGTCGCCAAC	705
Db	50759	YGYGYWSSSADBSYSSHYBGSYMCRKBSYASSIDVRYKBTGIRYVRSCKSWI	50818
Qy	706	TGCCAACACGACCTGGCATGGCTTGTTGGGGTCATCTAATTCTCCTATAAAGAATCAC	765
Db	50819	YVKYBSMSVSAKSKGHKVBRSSMCSRRKDMSTSMWRSMKDRKCSRCSHSHYSKSMGRK	50878
Qy	766	CTCCCATATTCTCGGCTCTGPGGGCACCTGTGCCATTCACTGGCACACGGTG	825
Db	50879	WTSDVDRCYACSSBKYKMRKKHFKKDSBHSWGRSHMKRGAWRTRSMRCSYHDSS	50938
Qy	826	CAGGCACACACCGGCATAAAAATAGGGCTCTTGCGGGCACGATCTGGACCGAGGGC	885
Db	50939	MVDKSRMRMGSAMRGMKCYCTSSMRSTSRSMKSSWRNGSKCYCYGWSCTMRSMG	50998
Qy	886	GCAATGGAATTCGAGGCTGTGATCAGGGTGTGGATCGGCCAGGGAAACCTGAAATGGCGC	945
Db	50999	CSYSTGGSSPKRCMKYCKNGRGRSSRYCKSRMSMAGGMSASMSMGKNTSGMSGRCCA	51058
Qy	946	GCATCAGTAGACCGGACGTACCTCTGAGAAAGGTCTGTC	989
Db	51059	RSGRMAYCSCRKCKSKRCKSCMRCKSYKCMRYGSWIMYKC	51102
RESULT 3			
AL77354	LOCUS	216878 bp DNA linear HTG 17-AUG-2002	
	DEFINITION	Mus musculus chromosome 2 clone RP23-183012, *** SEQUENCING IN PROGRESS	
		2 unordered pieces.	
	VERSION	AL773534	
	VERSION	AL773534.9	
	HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.		
	KEYWORDS		
	SOURCE	house mouse.	
	ORGANISM	Mus musculus	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
	REFERENCE	1 ('bases 1 to 216878)	
	AUTHORS	Johnson,C.	
	TITLE	Direct Submission	
	JOURNAL	Submitted (16 AUG-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail inquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk	
	COMMENT	On Aug 21, 2002 this sequence version replaced g1:22002783.	
	Center: Wellcome Trust Sanger Institute		
	Center project name:	bm183012	
	Project Information		
	-----	Summary Statistics	
	Assembly program:	XGAP4; version 4.5	
	Chemistry:	Dye-terminator; 100% of reads	
	Consensus quality:	216691 bases at least Q40	
	Consensus quality:	216754 bases at least Q30	
	Consensus quality:	216766 bases at least Q20	
	Insert size:	216778; sum-of-contigs	
	Insert size:	211070; 5.8x error; agarose-fp	
	Quality coverage:	7.01x in Q20 bases; sum-of-contigs Quality coverage: 7.25x in Q20 bases; agarose-fp	
	* NOTE:	This is a 'working draft' sequence. It currently	
	* consists of 2 contigs. The true order of the pieces		
	* is not known and their order in this sequence record is		
	* arbitrary. Gaps between the contigs are represented as		
	* runs of N, but the exact sizes of the gaps are unknown.		
	* This record will be updated with the finished sequence		
	* as soon as it is available and the accession number will		

* be preserved.

* 1 15291: contig of 15291 bp in length

* 15292 15391: gap of 100 bp

* 15392 216878: contig of 201487 bp in length.

FEATURES

Source	misc_feature	Query Match	Match	Score	DB	Length	Pr.M.
	/organism="Mus musculus"	Best Local Similarity	4.3%	44.4;	2;	216878	
	/db_xref="taxon:1090"	Local Similarity	47.5%	Pred. No.	0	31;	
	/chromosome="2"	Matches	132;	Mismatches	0;		
	/clone="RP23-183O12"	Conservative		Indels	146;		
	/clone_1b="RPCI-23"						
	1..15291						
	/note="assembly_fragment:02090"						
	fragment_chain.1"						
	15392..216878						
	/note="assembly_fragment:03034"						
	fragment_chain.1"						
	1..15291						

BASE COUNT

ORIGIN	70260	a	41471	c	41503	g	63544	t	100	others
--------	-------	---	-------	---	-------	---	-------	---	-----	--------

RESULT 4

ORGANISM	AF457660	AF457660	AF457660	AF457660	AF457660	AF457660	AF457660	AF457660	AF457660	AF457660
SOURCE	Castanea dentata	Bukayoya; Viridiplanteae; Streptophyta; Embryophyta; Tracheophytes; Rosidae; euroids I; Fagales; Fagaceae; Castanea.	Castanea dentata	Spermatophyta; Magnoliophyta; eudicots; core eudicots	clone	ACS2	vascular protein gene	linear	Pr.M.	
REFERENCE	Connors, B.J., Maynard, C.A., and Powell, W.A.									
AUTHORS	Connors, B.J., Maynard, C.A., and Powell, W.A.									
TITLE	Cloning and characterization of promoters from American chestnut reporter genes capable of directing reporter gene expression in transgenic Arabidopsis plants									
JOURNAL	Unpublished									
REFERENCE	2 (bases 1 to 685)									
AUTHORS	Connors, B.J., Maynard, C.A., and Powell, W.A.									
TITLE										
JOURNAL										

FEATURES

Source	
	/organism="Castanea dentata"
	/db_xref="taxon:134033"
	/clone="ACCS"

misc_feature 1. .685 /note="vascular protein promoter region"
 BASE COUNT 253 a 126 c 117 g 189 t
 ORIGIN

Query Match 4.2%; Score 42.8; DB 8; Length 685;
 Best Local Similarity 63.7%; Pred. No. 0.31;
 Matches 65; Conservative 0; Mismatches 37; Indels 0; Gaps 0;
 /product="MHC class I related protein 1"
 /db_xref="GI:4104807"

Qy 1 CTTACTATAGGGCACGCCGTTGGTCACGGCCGGCTGGTCTGAA 46
 Db 8 CTTACTATAGGGCACGCCGTTGGTCACGCCGGCTGGTCTGAA 60

Qy 61 GATGCATATCTGTGAGAGATAACATCACAAATTCAAGCT 102
 Db 68 AAGATAAAAAATATTATAACTTAAGATAAAATAAGCT 109

RESULT 5
 LOCUS AF039526
 DEFINITION Homo sapiens MHC class I related protein 1 (MRI) gene, partial cds.
 ACCESSION AF039526
 VERSION AF039526.1 GI:4104807
 KEYWORDS
 SOURCE Homo sapiens
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 1423)
 Riegent, P., Wanner, V. and Bahram, S.
 Genomics, isoforms, expression, and phylogeny of the MHC class
 I-related MRI gene
 J. Immunol. 161 (8), 4066-4077 (1998)
 MEDLINE 9845457
 PUBMED 9780177
 REFERENCE Riegent, P., Wanner, V., Hauptmann, G. and Bahram, S.
 AUTHORS Direct Submission
 TITLE Submitted (22-DIC-1997) EG, Basel Institute for Immunology,
 Grenzacherstrasse 487, Basel 4005, Switzerland
 FEATURES Location/Qualifiers
 1..1423
 source
 /organism="Homo sapiens"
 /db_xref="Taxon:9606"
 /chromosome="1"
 /map="1q25"
 1..>1423
 gene
 misc_feature
 1..1389
 /gene="MRI"
 /note="contains putative promoter and 5' UTR"
 <1390..>1423
 /gene="MRI"
 /product="MHC class I related protein 1"
 /evidence="not_experimental"
 CDS
 1390..>1423
 /gene="MRI"
 /codon_start=1
 /evidence="experimental"
 /product="MHC class I related protein 1"
 /protein_id="AAD02172.1"
 /db_xref="GI:4104809"
 /translation="MEELMFLPLL"

BASE COUNT 342 a 334 c 310 g 437 t
 ORIGIN

Query Match 4.1%; Score 42.4; DB 9; Length 1423;
 Best Local Similarity 97.7%; Pred. No. 0.47;
 Matches 43; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 /product="MHC class I related protein 1"
 /db_xref="GI:4104810"

Qy 1 CTTACTATAGGGCACGCCGTTGGTCACGCCGGCTGGTCTGAA 44
 Db 1 CTTACTATAGGGCACGCCGTTGGTCACGCCGGCTGGTCTGAA 60

Db 3 CTTACTATAGGGCACGCCGTTGGTCACGCCGGCTGGTCTGAA 46

RESULT 6
 LOCUS AR124194
 DEFINITION Sequence 1 from patent US 6171857.
 VERSION AR124194
 KEYWORDS Unknown
 SOURCE Unknown
 ORGANISM Unclassified
 REFERENCE 1 (bases 1 to 6078)
 AUTHORS Hendrickson, E.A.
 TITLE Leucine zipper protein, KARP-1 and methods of regulating DNA
 dependent protein kinase activity
 Patent: US 6171857-A 1 09-FAN-2001;
 JOURNAL Location/Qualifiers
 FEATURES Source
 /organism="unknown"
 BASE COUNT 1806 a 1301 c 1392 g 1543 t
 ORIGIN
 Query Match 4.1%; Score 42.4; DB 6; Length 6078;
 Best Local Similarity 97.7%; Pred. No. 0.62;
 Matches 43; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 /product="unknown"

Qy 1 CTTACTATAGGGCACGCCGTTGGTCACGCCGGCTGGTCTGAA 44
 Db 9 CTTACTATAGGGCACGCCGTTGGTCACGCCGGCTGGTCTGAA 52

Continuation (4 of 4) of PFMAL4P1 from base 300001 (AL034557 Plasmodium falciparum 3D
 /product="unknown")
 RESULT 7
 PFMAL4P1_3/c
 WPCOMMENT
 Sequence split into 4 fragments
 Fragment Name Begin End
 PFMAL4P1_0 1 110000
 PFMAL4P1_1 100001 210000
 PFMAL4P1_2 200001 310000
 PFMAL4P1_3 300001 392333
 Continuation (4 of 4) of PFMAL4P1 from base 300001 (AL034557 Plasmodium falciparum 3D
 /product="unknown")
 Query Match 4.1%; Score 41.8; DB 2; Length 92633;
 Best Local Similarity 65.6%; Pred. No. 1.6;
 Matches 61; Conservative 0; Mismatches 32; Indels 0; Gaps 0;
 /product="unknown"

Qy 1 CTTACTATAGGGCACGCCGTTGGTCACGCCGGCTGGTCTGAA 46
 Db 7315 CTTACTATAGGGCACGCCGGCTGGTCTGAA 60
 Continuation (3 of 4) of PFMAL4P1 from base 200001 (AL034557 Plasmodium falciparum 3D
 /product="unknown")
 Query Match 4.1%; Score 41.8; DB 2; Length 110000;
 Best Local Similarity 65.6%; Pred. No. 1.6;
 Matches 61; Conservative 0; Mismatches 32; Indels 0; Gaps 0;
 /product="unknown"

Qy 1 CTTACTATAGGGCACGCCGTTGGTCACGCCGGCTGGTCTGAA 46

	source
/gene="HVT1"	1..166703 /organism="Homo sapiens" /db_xref="taxon:9606"
5'UTR	/product="high affinity sulfate transporter" 1025..1089
CDS	/gene="HVT1" /product="high affinity sulfate transporter" 1090..>1094
	/protein_id="AAC4664_1" /db_xref="GI:3643824" /translation="MP" /codon_start=1
BASE COUNT	286 a 261 c 243 g 304 t
ORIGIN	
Query Match	4.0%
Best Local Similarity	95.5%
Matches	42; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY	1 CTTACTATAGGGCACCGGTTGTCACGGCCGGCTGGTATGAA 44
Db	63 CTTACTATAGGGCACGGGTGGTCAGGCCGGCTGGTCA 106
RESULT 12	
AL139378/c	166703 bp DNA linear PRI 02-DEC-2000
LOCUS	Human DNA sequence from clone RP11-271B5 on chromosome 13 Contains
DEFINITION	a gene for a protein similar to ribosomal protein S7, the FG9 (fibroblast growth factor 9 (gfa-activating factor)) gene, ESTs, STSS, GSS and CPG islands, complete sequence.
ACCESSION	AL139378
VERSION	AL139378.15 GI:89247510
KEYWORDS	HTG; CPG island; FGF9; growth factor; ribosomal protein.
SOURCE	human.
ORGANISM	Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
REFERENCE	1 (bases 1 to 166703)
AUTHORS	Blakey S.
TITLE	Direct Submission
JOURNAL	Submitted (18-JUL-2000) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT	On Jun 4, 2000 this sequence version replaced gi:8246895. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.
	The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information
	http://www.sanger.ac.uk/projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 13, constructed by the Sanger Centre Chromosome Mapping Group. Further information can be found at http://www.sanger.ac.uk/Hsp/Chr13
	RP11-271B5 is from the library RPCT-11.1 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see http://bacpac.med.buffalo.edu/
VECTOR:	phACE3.6
	This sequence is the entire insert of clone RP11-271B5 The true left end of clone RP11-28A24 is at 82919 in this sequence. The true right end of clone RP11-110K8 is at 72244 in this sequence.
FEATURES	Location/Qualifiers

misc_feature
 11255..11498
 note="match GSS: Em: AQ264373"
 misc_feature
 11281..11566
 note="match GSS: Em: AQ554409"
 repeat_region
 note="M9R5A repeat; matches 3..188 of consensus"
 repeat_region
 note="L1M4 repeat; matches 2891..3182 of consensus"
 repeat_region
 note="L1M4 repeat; matches 1..345 of consensus"
 repeat_region
 note="L1M4 repeat; matches 1511..1736 of consensus"
 repeat_region
 note="L1M4 repeat; matches 1260..1464 of consensus"
 repeat_region
 note="L1M4 repeat; matches 1..303 of consensus"
 repeat_region
 note="L1M4 repeat; matches 1..16517
 note="14 copies 3 mer tta 78% conserved"
 repeat_region
 note="L1M4 repeat; matches 1..296 of consensus"
 repeat_region
 note="L1M4 repeat; matches 6..269 of consensus"
 repeat_region
 note="L1M4 repeat; matches 49..143 of consensus"
 misc_feature
 19023..19447
 note="match GSS: Em: A0630744"
 misc_feature
 19036..19496
 note="match GSS: Em: A0630723"
 repeat_region
 note="M9R5A repeat; matches 1..138 of consensus"
 repeat_region
 note="L1M4 repeat; matches 1..181 of consensus"
 repeat_region
 note="L2 repeat; matches 2612..2750 of consensus"
 repeat_region
 note="L2 repeat; matches 2 mer tg 100% conserved"
 repeat_region
 note="AlusX repeat; matches 1..312 of consensus"
 repeat_region
 note="L2 repeat; matches 2662..2739 of consensus"
 repeat_region
 note="L2 repeat; matches 2 mer tg 100% conserved"
 repeat_region
 note="AlusX repeat; matches 1..494 of consensus"
 misc_feature
 26705..27376
 note="L2 repeat; matches 2597..2697 of consensus"
 note="match GSS: Em: A037994"
 repeat_region
 note="AlusX repeat; matches 1..295 of consensus"
 repeat_region
 note="L2 repeat; matches 2597..2697 of consensus"
 note="match GSS: Em: A023172"
 misc_feature
 28519..28685
 note="AlusX repeat; matches 1..27220
 note="2 copies 34 mer 100% conserved"
 repeat_region
 note="L2 repeat; matches 1..27220
 note="AlusX repeat; matches 1..295 of consensus"
 repeat_region
 note="L2 repeat; matches 1..27220
 note="AlusX repeat; matches 1..295 of consensus"
 misc_feature
 29089..29230
 note="L2 repeat; matches 2597..2697 of consensus"
 note="AlusX repeat; matches 1..295 of consensus"
 repeat_region
 note="L2 repeat; matches 1..295 of consensus"
 repeat_region
 note="L2 repeat; matches 1..295 of consensus"
 misc_feature
 31396..31605
 note="AlusX repeat; matches 1..297 of consensus"
 repeat_region
 note="L2 repeat; matches 1..303 of consensus"

```

source 1..156325 repeat_region 20049..20099
       /organism="Homo sapiens" repeat_region /note="MUTLI repeat: matches 92. .142 of consensus"
       /db_xref="taxon:9606" /note="AluSp repeat: matches 1. .275 of consensus"
       /map="p12" complement(2145..21455)
       /clone="RP1-134N8" /note="match: GSS: Em:A0033647"
       /clone.lib="RPCI-1" complement(2147..21876)
repeat_region 1145..1261 /note="match: GSS: Em:A0528323"
repeat_region 1278..1446 /note="AluSC repeat: matches 1. .309 of consensus"
repeat_region 1686..1846 /note="MIR repeat: matches 29. .179 of consensus"
repeat_region 2396..2495 /note="5 copies 20 mer 75% conserved"
repeat_region 2400..2485 /note="43 copies 2 mer ta 77% conserved"
repeat_region 3504..3563 /note="3 copies 20 mer 90% conserved"
repeat_region 4422..4603 /note="HAL1 repeat: matches 579. .771 of consensus"
repeat_region 4841..4897 /note="TLMEC repeat: matches 1318. .1368 of consensus"
repeat_region 4904..5263 /note="L1R18A repeat: matches 1. .603 of consensus"
repeat_region 5275..5520 /note="123 copies 2 mer tt 54% conserved"
repeat_region 6040..6834 /note="L1PA3 repeat: matches 5351. .6146 of consensus"
repeat_region 7290..7597 /note="AluX repeat: matches 1. .312 of consensus"
repeat_region 8305..8851 /note="match: GSS: Em:B66714"
repeat_region 8318..8819 /note="match: GSS: Em:AK228387"
repeat_region 8318..8680 /note="L1ME3 repeat: matches 5802. .6155 of consensus"
repeat_region 8809..8924 /note="MIR repeat: matches 119. .230 of consensus"
repeat_region 9121..9280 /note="MIR repeat: matches 45. .206 of consensus"
repeat_region 9296..9446 /note="MER5A repeat: matches 9. .180 of consensus"
repeat_region 10122..10887 /note="L1MBB8 repeat: matches 5379. .6171 of consensus"
repeat_region 10898..11043 /note="AluJb repeat: matches 159. .305 of consensus"
repeat_region 11097..11191 /note="L1MBB8 repeat: matches 5276. .5374 of consensus"
repeat_region 12638..12711 /note="37 copies 2 mer tt 67% conserved"
repeat_region 12921..13046 /note="FLAM_A repeat: matches 1. .127 of consensus"
repeat_region 14487..14958 /note="L2 repeat: matches 2225. .2749 of consensus"
repeat_region 15346..15405 /note="3 copies 20 mer 86% conserved"
repeat_region 15357..15404 /note="24 copies 2 mer ca 95% conserved"
repeat_region 15662..15798 /note="FLAM_A repeat: matches 1. .132 of consensus"
repeat_region 15809..15914 /note="MIR repeat: matches 74. .183 of consensus"
repeat_region 16324..16400 /note="MER5A repeat: matches 25. .106 of consensus"
repeat_region 17365..17492 /note="MER5B repeat: matches 2. .133 of consensus"
repeat_region 17481..17553 /note="MER5A repeat: matches 17. .100 of consensus"
repeat_region 17529..17978 /note="match: GSS: Em:AQ231626"
repeat_region 17852..18269 /note="MSIB repeat: matches 2. .426 of consensus"
repeat_region 24306..24383 /note="L2 repeat: matches 2347. .2426 of consensus"
repeat_region 24384..24681 /note="AluSp repeat: matches 1. .299 of consensus"
repeat_region 24682..24710 /note="L2 repeat: matches 2426. .2454 of consensus"
repeat_region 24711..25001 /note="AluX repeat: matches 4. .308 of consensus"
repeat_region 25002..25244 /note="L2 repeat: matches 2454. .2687 of consensus"
repeat_region 26228..26594 /note="TMEC repeat: matches 1. .371 of consensus"
repeat_region 26839..27151 /note="match: GSS: Em:AO084144"
repeat_region 27461..27498 /note="AluYb repeat: matches 1. .313 of consensus"
repeat_region 28252..28320 /note="L1ME repeat: matches 5266. .5266 of consensus"
repeat_region 31219..31288 /note="L1ME repeat: matches 159. .227 of consensus"
repeat_region 30530..30760 /note="match: SRS: Em:G24652"
repeat_region 30812..31214 /note="MUTLI repeat: matches 92. .385 of consensus"
repeat_region 31301..31374 /note="L1ME repeat: matches 5266. .5266 of consensus"
repeat_region 31562..31677 /note="35 copies 2 mer at 80% conserved"
repeat_region 31971..32384 /note="L1ME repeat: matches 274. .671 of consensus"
repeat_region 32220..32853 /note="L1ME repeat: matches 4988. .5266 of consensus"
repeat_region 32250..32300 /note="4 copies 20 mer 75% conserved"
repeat_region 32750..33158 /note="L1ME repeat: matches 4430. .4548 of consensus"
repeat_region 33194..33231 /note="19 copies 2 mer tg 85% conserved"
repeat_region 33387..34525 /note="7 copies 77 mer 95% conserved"
repeat_region 34715..35116 /note="match: GSS: Em:AK540824"
repeat_region 35471..35703 /note="match: GSS: Em:AK36937"
repeat_region 35743..36053 /note="MIR repeat: matches 21. .262 of consensus"
repeat_region 36568..36832 /note="AluY repeat: matches 1. .311 of consensus"
repeat_region 38453..38751 /note="AluJb repeat: matches 1. .254 of consensus"
repeat_region 38803..38939 /note="AluSc repeat: matches 1. .300 of consensus"

```

repeat_region /note="L2 repeat: matches 2577. .2710 of consensus" 39012. 39140
 misc_feature /note="MLTHI repeat: matches 339. .541 of consensus" complement:40115. .40665
 /note="match: GSS: Em:AQ424883"
 repeat_region /note="L2 repeat: matches 2605. .2750 of consensus"

Query Match 3.9%; Score 40.6; DB 9; Length 156325;
 Best Local Similarity 55.2%; Pred. No. 4;
 Matches 79; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 335 AGTTTCATCCATGCCAACAGAAATCCGCGAGAAGATTCTACACATAAAGGA 394
 Db 151285 AGTTTATTTTAAATTAACATTTATTTATTTGTGTTGTTAGAGA 151344
 gene /db_xref="Taxon:5722"
 796..1173

QY 395 GCACAGAGTTAACCTAACATCAATTAAACAGAGTTGCAATTAGACTTTAGAGA 454
 CDS /db_xref="GI:81132115"
 796..1173

Db 151345 TAATAAATGGTAAACTATAATCTAAAGAAATGTATCTCTTCAGTTCTAGGAA 151404
 /gene="Tvp14"
 /note="intracellular"
 /codon_start-1
 /product="Tvp14"
 /protein_id="AAF73223.1"
 /translation="MAFEFTTYDLLSRHEVIAEFGOLOHICPREFMTSLCPDRCDHAD
 VALEKVLNEYTKPEGEYGDPKHETIVDVKKKFVNODPKIQEYCXTLEVGGKYRCV
 DHUYLNRRNGSVPERPCTEVPL"

BASE COUNT 519 a 242. c 233 g 469 t
 ORIGIN

Query Match 3.9%; Score 40.4; DB 3; Length 1463;
 Best Local Similarity 97.6%; Pred. No. 1.9;
 Matches 41; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 TACATAGGCACCGTGTCGACGCCGGCTGGTGGATGAA 44
 Db 11 TACATAGGGCACCGTGGTCGACGCCGGCTGGTGGATGAA 52

Search completed: December 28, 2002, 01:04:17
 Job time : 3375 secs

RESULT 14
 LOCUS AX320057 598 bp DNA linear PAT 14-DEC-2001
 DEFINITION Sequence 14 from Patent WO0181606.
 ACCESSION AX320057
 VERSION AX320057.1 GI:17901561
 KEYWORD SOURCE
 ORGANISM Oryza sativa
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzeae; Oryza.
 REFERENCE 1
 AUTHORS Perera, J. R.; Iju, M. and Ray, A.
 TITLE Polynucleotide sequences from rice
 JOURNAL Patent: WO 0181606-A 14 01-NOV-2001;
 Akkadix Corporation (US)
 FEATURES Location/Qualifiers
 source 1. .598
 /organism="Oryza sativa"
 /db_xref="Taxon:4530"
 BASE COUNT 174 a 131 c 113 g 179 t 1 others
 ORIGIN

Query Match 3.9%; Score 40.4; DB 6; Length 598;
 Best Local Similarity 75.8%; Pred. No. 1.6;
 Matches 50; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 4 ACTATAGGCACCGTGTCGACGCCGGCTGGTGGATGAGGTGGAAACCTCAGTGGAT 63
 Db 1 ACTATAGGCACCGTGTCGACGCCGGCTGGTGGATCAGTGGTCAAACACTGTGCA 60
 QY 64 GCATAT 69
 Db 61 TAATAT 66

RESULT 15
 AF153014 AF153014 1463 bp DNA linear INV 10-APR-2002
 LOCUS AF153014 Trichomonas vaginalis Tvp14 (tvp14) gene, complete cds.
 DEFINITION AF153014
 ACCESSION AF153014.1 GI:8132114
 VERSION
 KEYWORDS SOURCE Trichomonas vaginalis.
 ORGANISM Trichomonas vaginalis; Parabasalidea; Trichomonadida; Trichomonidae;
 Eukaryota; Parabasalidea; Trichomonadida; Trichomonas; Trichomonadinae.

Gencore version 5.1.3
Copyright (C) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model
Run on: December 27, 2002, 20:39:18 ; Search time 55 Seconds
(without alignments)
5743.219 Million cell updates/sec

Title: US-09-945-376-3
Perfect score: 1030
Sequence: cttaataggcacgcgtg.....tgttatatacacaaggctcg 1030

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA:
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq;*: /cgn2_6/ptodata/1/ina/5B_COMB.seq;*: /cgn2_6/ptodata/1/ina/6A_COMB.seq;*: /cgn2_6/ptodata/1/ina/6B_COMB.seq;*: /cgn2_6/ptodata/1/ina/PCUS_COMB.seq;*: /cgn2_6/ptodata/1/ina/backfiles1.seq;*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match	Length	DB	ID	Description	
1	42.4	4.1	6078	4	US-09-173-914-1	Sequence 1, Appli		
2	39.8	3.9	3718	4	US-09-424-283-6	Sequence 6, Appli		
3	39.6	3.8	7218	1	US-08-232-463-14	Sequence 14, Appli		
4	39	3.8	341	4	US-09-43-195A-1	Sequence 1, Appli		
5	39	3.8	4526	4	US-09-424-283-7	Sequence 7, Appli		
c	38.4	3.7	921	4	US-09-377-648-4	Sequence 4, Appli		
7	38.4	3.7	2791	4	US-09-570-367C-1	Sequence 1, Appli		
c	38	3.7	336	4	US-09-276-599-13	Sequence 13, Appli		
c	38	3.7	763	4	US-09-276-599-14	Sequence 14, Appli		
c	10	37.8	565	4	US-09-323-195A-5	Sequence 5, Appli		
c	11	37.8	3.7	1924	4	US-09-424-283-5	Sequence 5, Appli	
c	12	37.4	3.6	478	4	US-08-913-014A-8	Sequence 18, Appli	
c	13	37.4	3.6	2791	4	US-09-570-367C-1	Sequence 1, Appli	
c	14	37.2	3.6	1478	4	US-09-545-814-28	Sequence 28, Appli	
c	15	37.2	3.6	1478	4	US-09-545-814-10	Sequence 30, Appli	
c	16	37.2	3.6	1677	4	US-09-545-814-13	Sequence 13, Appli	
c	17	37.2	3.6	1677	4	US-09-545-814-15	Sequence 15, Appli	
c	18	37.2	3.6	1749	4	US-09-545-814-4	Sequence 4, Appli	
c	19	37.2	3.6	1749	4	US-09-545-814-6	Sequence 6, Appli	
c	20	37.2	3.6	1919	4	US-09-545-814-31	Sequence 31, Appli	
c	21	37.2	3.6	1919	4	US-09-545-814-33	Sequence 33, Appli	
c	22	37.2	3.6	2610	4	US-09-545-814-3	Sequence 1, Appli	
c	23	37.2	3.6	2610	4	US-09-545-814-3	Sequence 3, Appli	
c	24	37	3.6	340	4	US-09-323-195A-3	Sequence 5, Appli	
c	25	36	3.5	255	4	US-09-257-195A-5	Sequence 6, Appli	
c	26	36	3.5	342	4	US-09-323-195A-6	Sequence 37, Appli	
c	27	36	3.5	555	4	US-08-905-223-37		

ALIGNMENTS

RESULT 1									
US-09-173-914-1									
Sequence 1, Appli									
Sequence 2, Appli									
Sequence 3, Appli									
Sequence 4, Appli									
Sequence 5, Appli									
Sequence 6, Appli									
Sequence 7, Appli									
Sequence 8, Appli									
Sequence 9, Appli									
Sequence 10, Appli									
Sequence 11, Appli									
Sequence 12, Appli									
Sequence 13, Appli									
Sequence 14, Appli									
Sequence 15, Appli									
Sequence 16, Appli									
Sequence 17, Appli									
Sequence 18, Appli									
Sequence 19, Appli									
Sequence 20, Appli									
Sequence 21, Appli									
Sequence 22, Appli									
Sequence 23, Appli									
Sequence 24, Appli									
Sequence 25, Appli									
Sequence 26, Appli									
Sequence 27, Appli									
Sequence 28, Appli									
Sequence 29, Appli									
Sequence 30, Appli									
Sequence 31, Appli									
Sequence 32, Appli									
Sequence 33, Appli									
Sequence 34, Appli									
Sequence 35, Appli									
Sequence 36, Appli									
Sequence 37, Appli									
Sequence 38, Appli									
Sequence 39, Appli									
Sequence 40, Appli									
Sequence 41, Appli									
Sequence 42, Appli									
Sequence 43, Appli									
Sequence 44, Appli									
Sequence 45, Appli									
Sequence 46, Appli									
Sequence 47, Appli									
Sequence 48, Appli									
Sequence 49, Appli									
Sequence 50, Appli									
Sequence 51, Appli									
Sequence 52, Appli									
Sequence 53, Appli									
Sequence 54, Appli									
Sequence 55, Appli									
Sequence 56, Appli									
Sequence 57, Appli									
Sequence 58, Appli									
Sequence 59, Appli									
Sequence 60, Appli									
Sequence 61, Appli									
Sequence 62, Appli									
Sequence 63, Appli									
Sequence 64, Appli									
Sequence 65, Appli									
Sequence 66, Appli									
Sequence 67, Appli									
Sequence 68, Appli									
Sequence 69, Appli									
Sequence 70, Appli									
Sequence 71, Appli									
Sequence 72, Appli									
Sequence 73, Appli									
Sequence 74, Appli									
Sequence 75, Appli									
Sequence 76, Appli									
Sequence 77, Appli									

; PRIOR APPLICATION NUMBER: PCT/US98/10465
 ; PRIOR FILING DATE: 1998-05-21
 ; PRIOR APPLICATION NUMBER: US 60/047,568
 ; PRIOR FILING DATE: 1997-05-22
 ; NUMBER OF SEQ ID NOS: 15
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 7
 ; LENGTH: 4526
 ; TYPE: DNA
 ; ORGANISM: Glycine max
 ; US-09-424-283-7

Query Match 3.8%; Score 39; DB 4; Length 4526;
 Best Local Similarity 100.0%; Pred. No. 0.042;
 Matches 39; Conservative 0; Mismatches 0;
 Indels 0; Gaps 0;

Qy 2 TTACTATAGGCACGGCAGTCGACGGCCGGCTGGTA 40
 Db 97 TTACTATAGGCACGGCAGTCGACGGCCGGCTGGTA 135

RESULT 6
 ; Sequence 4, Application US/09377648
 ; Patent No. 6225549
 ; GENERAL INFORMATION:
 ; APPLICANT: Lappergard, Kathryn
 ; TITLE OF INVENTION: Seed-preferred Promoters
 ; FILE REFERENCE: 0869
 ; CURRENT APPLICATION NUMBER: US/09/377,648
 ; CURRENT FILING DATE: 1999-08-19
 ; EARLIER APPLICATION NUMBER: US 60/097,233
 ; EARLIER FILING DATE: 1998-08-20
 ; NUMBER OF SEQ ID NOS: 9
 ; SOFTWARE: FastSEQ for Windows Version 3.0
 ; SEQ ID NO 4
 ; LENGTH: 921
 ; TYPE: DNA
 ; ORGANISM: Zea mays
 ; FEATURE:
 ; NAME/KEY: promoter
 ; LOCATION: (1)...(922)
 ; US-09-377-648-4

Query Match 3.7%; Score 38.4; DB 4; Length 921;
 Best Local Similarity 97.5%; Pred. No. 0.028;
 Matches 39; Conservative 0; Mismatches 1;
 Indels 0; Gaps 0;

Qy 1 CTTACTATAGGCACGGCAGTCGACGGCCGGCTGGTA 40
 Db 912 CTCACTATAGGCACGGCAGTCGACGGCCGGCTGGTA 873

RESULT 7
 ; Sequence 1, Application US/09570367C
 ; Patent No. 6338851
 ; GENERAL INFORMATION:
 ; APPLICANT: Gorczynski, Reginald M.
 ; TITLE OF INVENTION: Methods and Compositions for Immunomodulation
 ; FILE REFERENCE: 9579-1
 ; CURRENT APPLICATION NUMBER: US/09/570,367C
 ; CURRENT FILING DATE: 2000-05-05
 ; PRIOR APPLICATION NUMBER: US 60/064,764
 ; PRIOR FILING DATE: 1997-11-07
 ; NUMBER OF SEQ ID NOS: 22
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 1
 ; LENGTH: 2791
 ; TYPE: DNA
 ; ORGANISM: Mus musculus
 ; US-09-570-367C-1

Query Match 3.7%; Score 38; DB 4; Length 2791;
 Best Local Similarity 87.5%; Pred. No. 0.051;
 Matches 42; Conservative 0; Mismatches 6;
 Indels 0; Gaps 0;

Qy 4 ACTATAGGGCACGGCAGTCGACGGCCGGCTGGTATGAGGTGGAA 51
 Db 1 ACTATAGGGCACGGCAGTCGACGGCCGGCTGGTATGAGGTGGAA 48

RESULT 8
 ; Sequence 13, Application US/09276599
 ; Patent No. 6380459
 ; GENERAL INFORMATION:
 ; APPLICANT: Perera, J. Ranjan
 ; TITLE OF INVENTION: Composition and methods for the
 ; TITLE OF INVENTION: Modification of gene expression
 ; FILE REFERENCE: 11000.1036
 ; CURRENT APPLICATION NUMBER: US/09/276,599
 ; CURRENT FILING DATE: 1999-03-25
 ; NUMBER OF SEQ ID NOS: 21
 ; SOFTWARE: FastSEQ for Windows Version 3.0
 ; SEQ ID NO 13
 ; LENGTH: 336
 ; TYPE: DNA
 ; ORGANISM: Pinus radiata
 ; US-09-276-599-13

Query Match 3.7%; Score 38; DB 4; Length 336;
 Best Local Similarity 100.0%; Pred. No. 0.022;
 Matches 38; Conservative 0; Mismatches 0;
 Indels 0; Gaps 0;

Qy 2 TTACTATAGGCACGGCAGTCGACGGCCGGCTGGGT 39
 Db 334 TTACTATAGGCACGGCAGTCGACGGCCGGCTGGT 297

RESULT 9
 ; Sequence 14, Application US/09276599
 ; Patent No. 6380459
 ; GENERAL INFORMATION:
 ; APPLICANT: Perera, J. Ranjan
 ; TITLE OF INVENTION: Composition and methods for the
 ; TITLE OF INVENTION: Modification of gene expression
 ; FILE REFERENCE: 11000.1036
 ; CURRENT APPLICATION NUMBER: US/09/276,599
 ; CURRENT FILING DATE: 1999-03-25
 ; NUMBER OF SEQ ID NOS: 21
 ; SOFTWARE: FastSEQ for Windows Version 3.0
 ; SEQ ID NO 14
 ; LENGTH: 763
 ; TYPE: DNA
 ; ORGANISM: Pinus radiata
 ; US-09-276-599-14

Query Match 3.7%; Score 38; DB 4; Length 763;
 Best Local Similarity 100.0%; Pred. No. 0.034;
 Matches 38; Conservative 0; Mismatches 0;
 Indels 0; Gaps 0;

Qy 2 TTACTATAGGGCACGGCTGGTACGGCGCCGGCTGGT 39
 Db 761 TTACTATAGGGCACGGCTGGTACGGCGCCGGCTGGT 724

RESULT 10
 ; Sequence 5, Application US/09323195A
 ; Patent No. 6462257
 ; GENERAL INFORMATION:

```

; APPLICANT: Pullman, Gerald
; APPLICANT: Cairney, John
; APPLICANT: Pereira, Ranjan
; TITLE OF INVENTION: VICILIN-LIKE SEED STORAGE PROTEIN GENE PROMOTER AND METHODS OF USING THE SAME
; FILE REFERENCE: IPST0009
; CURRENT APPLICATION NUMBER: US/09/323,195A
; CURRENT FILING DATE: 1990-06-01
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 5
LENGTH: 565
TYPE: DNA
ORGANISM: Pinus taeda
FEATURE:
; OTHER INFORMATION: n at 489 is a, c, g, or t
; OTHER INFORMATION: n at 503 is a, c, g, or t
; OTHER INFORMATION: n at 504 is a, c, g, or t
; OTHER INFORMATION: n at 522 is a, c, g, or t
; OTHER INFORMATION: n at 533 is a, c, g, or t
; OTHER INFORMATION: n at 543 is a, c, g, or t
; OTHER INFORMATION: n at 549 is a, c, g, or t
; OTHER INFORMATION: n at 564 is a, c, g, or t
US-09-323-195A-5

Query Match Score 37.8; DB 4; Length 565;
Best Local Similarity 95.1%; Pred: No. 0, 0.033; Indels 0; Gaps 0;
Matches 39; Conservative 0; Mismatches 2;

Qy 4 ACTATAGGCCACGGTGTGCGACGGCCGGCGCTGATGAA 44
Db 1 ACTATAGGCCACGGTGTGCGACGGCCGGCGCTGATGAA 41

```

RESULT 11
US-09-424 283-5/c
; Sequence 5, Application US/09424283
; Patent No. 6437219
; GENERAL INFORMATION:
; APPLICANT: Grimes, et al.
; TITLE OF INVENTION: Sucrose binding proteins
; FILE REFERENCE: 4630-50206
; CURRENT APPLICATION NUMBER: US/09/424,283
; CURRENT FILING DATE: 1999-11-19
; PRIOR APPLICATION NUMBER: PCT/US98/10465
; PRIOR FILING DATE: 1998-05-21
; PRIOR APPLICATION NUMBER: US 60/047,568
; PRIOR FILING DATE: 1997-05-22
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0

RESULT 12
US-08-913-014-A18
; Sequence 18, Application US/08913014A
; Patent No. 6235878
; GENERAL INFORMATION.

Best Local Similarity 97.4%; Pred. No. 0.11; Matches 38; Conservative 0; Mismatches 1; Indels 0; Gaps 0; LENGTH: 1478; TYPE: DNA ; ORGANISM: Tagged Ctenocephalides felis us-09-545-814-30

Query Match Score 37.2; DB 4; Length 1478; Best Local Similarity 47.1%; Pred. No. 0.087; Mismatches 0; Indels 0; Gaps 0; Matches 114; Conservative 0;

Qy 66 ATATACTGCTGAGAGATAACATCACATTCAAGCTTCATGTCACTACAGCCTACAGC 125
Db 1023 AGACATTAACGGTGGATACTCGAAAGTTCTGGTACATGATATCCCTCTG 1082

Qy 126 CATTCCGAGGAGGGCTTAGGACGGCCGTAAATCGAACATGCCAAAGAATG 185
Db 1083 CAGGCCAAATTGGCATGGACGCCATGGTATCCAGTTGCCCTGTTGTTGTT 1142

Qy 186 GATTAAACAGCTCTATCTAGGCAGCTAAACGCAAGCTAAATCGTATAGTCATTTTATC 245
Db 1143 GAAAGCCCTTCGCAATTCTGGACAAAGATAAAATTGTCCTTGCGAAAACGTGCC 1202

Qy 246 TCCCCAACGGCATPTGTAAGCAGCTAAATCGATSTAATGTCIAC 305
Db 1203 ACCCTCATCACGCCACGGATACTCCAAATCCAAAGTTCGAAACCGTCAAACGTGATTC 1262

Qy 306 TT 307
Db 1263 GT 1264

Search completed: December 28, 2002, 00:01:55
Job time : 63 secs

Best Local Similarity 47.1%; Pred. No. 0.087; Mismatches 128; Indels 0; Gaps 0;

Qy 66 ATATACTGCTGAGAGATAACATCACATTCAAGCTTCATGTCACTACAGCCTACAGC 125
Db 456 AGACATTAGCTGGATAGCTCPATAATTGCAAAAGTTCTGGTACATGATATCCCTCTG 397

Qy 126 CATTCCGAGGAGGGCTTAGGACGGCCGTAAATCGAACATGCCAAAGAATG 185
Db 396 CAGCTAAATTGGCAATGGCACGGCATGGTATCTCCAGTTTGCTTGTGTTGTT 337

Qy 186 GATTAAACAGCTCTATCGAGGACAGCTAAACGCAAGCTTGTATAGTCATTTTATC 245
Db 336 GAAAGCCCTTCGCAATTCTGGACAAAGATAAAATTGTCCTTGCGAAAACGTGCC 277

Qy 246 TCCCCAACGGCATPTGTAAGCAGCTAAATCGATSTAATGTCIAC 305
Db 276 ACCCTCATCACGCCACGGATACTCCAAATCGAAACCGTCAAACGTGATTC 217

Qy 306 TT 307
Db 216 GT 215

RESULT 155
US-09-545-814-30
; Sequence 30, Application US/09545814
; Patient No. 6416977
; GENERAL INFORMATION:
; APPLICANT: Becher, Anna M.
; TITLE OF INVENTION: FLEA CHITINASE NUCLEIC ACID MOLECULES, PROTEINS AND
; FILE REFERENCE: FC-5-C1
; CURRENT APPLICATION NUMBER: US/09-545-814
; CURRENT FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: 60/128, 833
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 30

Result No.	Score	Query Match Length DB ID	Description
C 1	45	4 . 4	BH747511 SALK_0180
C 2	43 . 6	4 . 2	BH251364 SALK_0114
C 3	43 . 4	4 . 2	BH617520 SALK_0372
C 4	42 . 8	4 . 2	BH613783 SALK_0349
C 5	42 . 8	4 . 2	BH250931 SALK_0107
C 6	42 . 8	4 . 2	BH634101 SALK_0447

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length DB ID	Description
17:	gb_gss:*	*	BH747511 SALK_0180
18:	en_gss_hum:*	*	BH251364 SALK_0114
19:	en_gss_inv:*	*	BH617520 SALK_0372
20:	en_gss_pln:*	*	BH613783 SALK_0349
21:	en_gss_vrt:*	*	BH250931 SALK_0107
22:	en_gss_frn:*	*	BH634101 SALK_0447
23:	en_gss_mam:*	*	
24:	en_gss_mus:*	*	
25:	en_gss_other:*	*	
26:	en_gss_pro:*	*	
27:	en_gss_rnd:*	*	
15:	en_esttun:*	*	
16:	em_estcom:*	*	
17:	gb_htc:*	*	
11:	gb_htc:*	*	
12:	gb_est3:*	*	
13:	gb_est4:*	*	
14:	gb_est5:*	*	
15:	gb_esttun:*	*	
16:	em_estcom:*	*	
17:	gb_gss:*	*	
18:	en_gss_hum:*	*	
19:	en_gss_inv:*	*	
20:	en_gss_pln:*	*	
21:	en_gss_vrt:*	*	
22:	en_gss_frn:*	*	
23:	en_gss_mam:*	*	
24:	en_gss_mus:*	*	
25:	en_gss_other:*	*	
26:	en_gss_pro:*	*	
27:	en_gss_rnd:*	*	

SUMMARY

8

JOURNAL
COMMENT

Unpublished (2001)
Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGNAL)

The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA

TeL: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salk.edu

This is single pass sequence recovered from the left border of

RESULT 1	BH747511/c	LOCUS	BH747511 388 bp DNA linear GSS 27-FBB-2002
		DEFINITION	SALK_018023 .45.85 .x Arabidopsis thaliana tDNA insertion lines
			Arabidopsis thaliana genomic clone SALK_018023 .45.85 .x , DNA sequence.
		ACCESSION	BH747511
		VERSION	1 (bases 1 to 388)
		KEYWORDS	GSS
		SOURCE	thale cress.
		ORGANISM	Arabidopsis thaliana; Streptophyta; Embryophyta; Tracheophyta; Eukaryota; Viridiplantae; Magnoliophyta; eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
		REFERENCE	1 Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P., Zimmerman,J., and Ecker,J.R.
		AUTHORS	
		TITLE	A Sequence-Indexed Library of Insertion Mutations in the Arabidopsis Genome
		JOURNAL	Unpublished (2001)
		COMMENT	Contact: Joseph R. Ecker

Scoring table: IDENTITY_NUC	Gapop 10.0 , Gapext 1.0	C 7	42 . 6	4 . 1	440	17	BH748887 SALK_0469
Perfect score: 1030	Sequence: 1 cttaatagggcacgcgtg.tgtataataccaaaggctctgg 1030	C 8	42 . 2	4 . 1	432	17	BH610787 SALK_0180
Run on: December 28, 2002, 00:00:48 ; Search time 1964 Seconds (without alignments)	8493.560 Million cell updates/sec	C 9	42 . 2	4 . 1	500	17	BH214178 SALK_0103
Title: US-09-945-376-3		C 10	42 . 2	4 . 1	1101	17	CNS001GU
Post-processing: Minimum Match 0% Maximum Match 100%	Listing first 45 summaries	C 11	42	4 . 1	129	17	BH688812 SALK_0002
Total number of hits satisfying chosen parameters: 32308132		C 12	42	4 . 1	220	17	BH251572 SALK_0118
Minimum DB seq length: 0		C 13	42	4 . 1	452	17	BH212774 SALK_0080
Maximum DB seq length: 2000000000		C 14	41 . 8	4 . 1	416	17	BH749138 SALK_0473
Post-processing: Minimum Match 0% Maximum Match 100%		C 15	41 . 6	4 . 0	457	17	BH72739 SALK_0061
Scoring table: IDENTITY_NUC	Gapop 10.0 , Gapext 1.0	C 16	41 . 6	4 . 0	471	17	BH251074 SALK_0109
Perfect score: 1030	Sequence: 1 cttaatagggcacgcgtg.tgtataataccaaaggctctgg 1030	C 17	41 . 6	4 . 0	502	17	BH253202 SALK_0146
Run on: December 28, 2002, 00:00:48 ; Search time 1964 Seconds (without alignments)	8493.560 Million cell updates/sec	C 18	41 . 4	4 . 0	120	17	BH750440 SALK_0048
Title: US-09-945-376-3		C 19	41 . 4	4 . 0	430	17	BH612486 SALK_0328
Post-processing: Minimum Match 0% Maximum Match 100%	Listing first 45 summaries	C 20	41 . 4	4 . 0	445	17	BH612500 SALK_0326
Total number of hits satisfying chosen parameters: 32308132		C 21	41 . 4	4 . 0	456	17	BH617406 SALK_0364
Minimum DB seq length: 0		C 22	41 . 4	4 . 0	456	17	BH633739 SALK_0431
Maximum DB seq length: 2000000000		C 23	41 . 2	4 . 0	404	17	BH617295 SALK_0362
Post-processing: Minimum Match 0% Maximum Match 100%	Listing first 45 summaries	C 24	41 . 2	4 . 0	465	17	BH70645 SALK_0031
Scoring table: IDENTITY_NUC	Gapop 10.0 , Gapext 1.0	C 25	41 . 2	4 . 0	469	17	BH749118 SALK_0473
Perfect score: 1030	Sequence: 1 cttaatagggcacgcgtg.tgtataataccaaaggctctgg 1030	C 26	41 . 2	4 . 0	235	17	BH251060 SALK_0109
Run on: December 28, 2002, 00:00:48 ; Search time 1964 Seconds (without alignments)	8493.560 Million cell updates/sec	C 27	41	4 . 0	248	17	BH254262 SALK_0162
Title: US-09-945-376-3		C 28	41	4 . 0	128	17	BH717573 SALK_0045
Post-processing: Minimum Match 0% Maximum Match 100%	Listing first 45 summaries	C 29	40 . 8	4 . 0	433	17	BH610810 SALK_0180
Scoring table: IDENTITY_NUC	Gapop 10.0 , Gapext 1.0	C 30	40 . 8	4 . 0	451	17	BH254421 SALK_0165
Perfect score: 1030	Sequence: 1 cttaatagggcacgcgtg.tgtataataccaaaggctctgg 1030	C 31	40 . 8	4 . 0	477	17	BH70269 SALK_0026
Run on: December 28, 2002, 00:00:48 ; Search time 1964 Seconds (without alignments)	8493.560 Million cell updates/sec	C 32	40 . 6	3 . 9	105	17	BH752334 SALK_0140
Title: US-09-945-376-3		C 33	40 . 6	3 . 9	121	17	BH616998 SALK_0358
Scoring table: IDENTITY_NUC	Gapop 10.0 , Gapext 1.0	C 34	40 . 6	3 . 9	128	17	BH752369 SALK_0184
Perfect score: 1030	Sequence: 1 cttaatagggcacgcgtg.tgtataataccaaaggctctgg 1030	C 35	40 . 6	3 . 9	218	17	BH170998 SALK_0035
Run on: December 28, 2002, 00:00:48 ; Search time 1964 Seconds (without alignments)	8493.560 Million cell updates/sec	C 36	40 . 6	3 . 9	266	17	BH611050 SALK_0297
Title: US-09-945-376-3		C 37	40 . 6	3 . 9	454	17	BH611042 SALK_0135
Scoring table: IDENTITY_NUC	Gapop 10.0 , Gapext 1.0	C 38	40 . 6	3 . 9	456	17	BH252514 SALK_0135
Perfect score: 1030	Sequence: 1 cttaatagggcacgcgtg.tgtataataccaaaggctctgg 1030	C 39	40 . 6	3 . 9	492	17	BH611714 SALK_0315
Run on: December 28, 2002, 00:00:48 ; Search time 1964 Seconds (without alignments)	8493.560 Million cell updates/sec	C 40	40 . 4	3 . 9	438	17	BH251902 SALK_0123
Title: US-09-945-376-3		C 41	40 . 4	3 . 9	460	17	BH188842 036_P_10
Scoring table: IDENTITY_NUC	Gapop 10.0 , Gapext 1.0	C 42	40 . 4	3 . 9	460	17	AL625781 T3 end of CNS07SRN
Perfect score: 1030	Sequence: 1 cttaatagggcacgcgtg.tgtataataccaaaggctctgg 1030	C 43	40 . 4	3 . 9	470	17	BH172612 SALK_0059
Run on: December 28, 2002, 00:00:48 ; Search time 1964 Seconds (without alignments)	8493.560 Million cell updates/sec	C 44	40 . 2	3 . 9	114	17	BH250920 SALK_0107
Title: US-09-945-376-3		C 45	40 . 2	3 . 9	116	17	BH251621 SALK_0118

Scoring table: IDENTITY_NUC	Gapop 10.0 , Gapext 1.0	C 7	42 . 6	4 . 1	440	17	BH748887 SALK_0469
Perfect score: 1030	Sequence: 1 cttaatagggcacgcgtg.tgtataataccaaaggctctgg 1030	C 8	42 . 2	4 . 1	500	17	BH214178 SALK_0103
Run on: December 28, 2002, 00:00:48 ; Search time 1964 Seconds (without alignments)	8493.560 Million cell updates/sec	C 9	42 . 2	4 . 1	1101	17	CNS001GU
Title: US-09-945-376-3		C 10	42 . 2	4 . 1	129	17	BH688812 SALK_0002
Post-processing: Minimum Match 0% Maximum Match 100%	Listing first 45 summaries	C 11	42	4 . 1	220	17	BH251572 SALK_0118
Scoring table: IDENTITY_NUC	Gapop 10.0 , Gapext 1.0	C 12	42	4 . 1	452	17	BH212774 SALK_0080
Perfect score: 1030	Sequence: 1 cttaatagggcacgcgtg.tgtataataccaaaggctctgg 1030	C 13	42	4 . 1	457	17	BH749138 SALK_0473
Run on: December 28, 2002, 00:00:48 ; Search time 1964 Seconds (without alignments)	8493.560 Million cell updates/sec	C 14	41 . 8	4 . 1	416	17	BH72739 SALK_0061
Title: US-09-945-376-3		C 15	41 . 6	4 . 0	457	17	BH747511 388 bp DNA linear GSS 27-FBB-2002
Scoring table: IDENTITY_NUC	Gapop 10.0 , Gapext 1.0	C 16	41 . 6	4 . 0	471	17	BH253202 SALK_0146
Perfect score: 1030	Sequence: 1 cttaatagggcacgcgtg.tgtataataccaaaggctctgg 1030	C 17	41 . 6	4 . 0	502	17	BH617295 SALK_0116
Run on: December 28, 2002, 00:00:48 ; Search time 1964 Seconds (without alignments)	8493.560 Million cell updates/sec	C 18	41 . 4	4 . 0	120	17	BH750440 SALK_0048
Title: US-09-945-376-3		C 19	41 . 4	4 . 0	430	17	BH612486 SALK_0328
Scoring table: IDENTITY_NUC	Gapop 10.0 , Gapext 1.0	C 20	41 . 4	4 . 0	445	17	BH612500 SALK_0326
Perfect score: 1030	Sequence: 1 cttaatagggcacgcgtg.tgtataataccaaaggctctgg 1030	C 21	41 . 4	4 . 0	456	17	BH617406 SALK_0364
Run on: December 28, 2002, 00:00:48 ; Search time 1964 Seconds (without alignments)	8493.560 Million cell updates/sec	C 22	41 . 4	4 . 0	456	17	BH633739 SALK_0431
Title: US-09-945-376-3		C 23	41 . 2	4 . 0	404	17	BH617295 SALK_0362
Scoring table: IDENTITY_NUC	Gapop 10.0 , Gapext 1.0	C 24	41 . 2	4 . 0	465	17	BH70645 SALK_0031
Perfect score: 1030	Sequence: 1 cttaatagggcacgcgtg.tgtataataccaaaggctctgg 1030	C 25	41 . 2	4 . 0	469	17	BH749118 SALK_0473
Run on: December 28, 2002, 00:00:48 ; Search time 1964 Seconds (without alignments)	8493.560 Million cell updates/sec	C 26	41 . 2	4 . 0	469	17	BH251060 SALK_0109
Title: US-09-945-376-3		C 27	41	4 . 0	235	17	BH254262 SALK_0162
Scoring table: IDENTITY_NUC	Gapop 10.0 , Gapext 1.0	C 28	41	4 . 0	456	17	BH717573 SALK_0045
Perfect score: 1030	Sequence: 1 cttaatagggcacgcgtg.tgtataataccaaaggctctgg 1030	C 29	40 . 8	4 . 0	433	17	BH610810 SALK_0180
Run on: December 28, 2002, 00:00:48 ; Search time 1964 Seconds (without alignments)	8493.560 Million cell updates/sec	C 30	40 . 8	4 . 0	451	17	BH254421 SALK_0165
Title: US-09-945-376-3		C 31	40 . 8	4 . 0	477	17	BH70269 SALK_0026
Scoring table: IDENTITY_NUC	Gapop 10.0 , Gapext 1.0	C 32	40 . 6	3 . 9	105	17	BH752334 SALK_0140
Perfect score: 1030	Sequence: 1 cttaatagggcacgcgtg.tgtataataccaaaggctctgg 1030	C 33	40 . 6	3 . 9	121	17	BH616998 SALK_0358
Run on: December 28, 2002, 00:00:48 ; Search time 1964 Seconds (without alignments)	8493.560 Million cell updates/sec	C 34	40 . 6	3 . 9	128	17	BH752369 SALK_0184
Title: US-09-945-376-3		C 35	40 . 6	3 . 9	218	17	BH170998 SALK_0035
Scoring table: IDENTITY_NUC	Gapop 10.0 , Gapext 1.0	C 36	40 . 6	3 . 9	266	17	BH611050 SALK_0297
Perfect score: 1030	Sequence: 1 cttaatagggcacgcgtg.tgtataataccaaaggctctgg 1030	C 37	40 . 6	3 . 9	454	17	BH611042 SALK_0135
Run on: December 28, 2002, 00:00:48 ; Search time 1964 Seconds (without alignments)	8493.560 Million cell updates/sec	C 38	40 . 6	3 . 9	456	17	BH252514 SALK_0135
Title: US-09-945-376-3		C 39	40 . 6	3 . 9	492	17	BH611714 SALK_0315
Scoring table: IDENTITY_NUC	Gapop 10.0 , Gapext 1.0	C 40	40 . 4	3 . 9	438</		

TDNA. This sequence lies within 300 bases of the 3' end of At1g57820.
 Class: TDNA tagged.

FEATURES source
 1. 388
 /organism="Arabidopsis thaliana"
 /strain="Columbia 0"
 /db_xref="taxon:3702"
 /clone="SALK_018023_45.85.x"
 /note="PCR was performed on Arabidopsis thaliana TDNA insertion lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at <http://signal.salk.edu/tDNA-protocols.html>"
 3 others

BASE COUNT 98 a 95 c 82 g 110 t
 ORIGIN 3 others

Query Match 4.43; Score 45; DB 17; Length 388;
 Best Local Similarity 78.33; Pred. No. 0.0094; Indels 0; Gaps 0;
 Matches 54; Conservative 0; Mismatches 15;

Qy 1 CTTACTATAGGGCACCGCTGGTCGAGGCCGGCTGGTATGAAGTGGAACCTCACTG 60
 Db 265 CTCACATAGGGCACCGGTGGTCGAGGCCGGCTGGTCAATTCAATGGAACAGGGAA 206

Qy 61 GATGCCAT 69
 Db 205 AATGCCAT 197

RESULT 2
 LOCUS BH251364/c 469 bp DNA linear GSS 28-NOV-2001
 DEFINITION SALK_011461 Arabidopsis thaliana TDNA insertion lines Arabidopsis thaliana genomic clone SALK_011461, DNA sequence.
 ACCESION BH251364
 VERSION BH251364.1 GI:17138342
 KEYWORDS GSS,
 SOURCE thale cress.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 REFERENCES Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P., Zimmerman,J., and Ecker,J.R.
 TITLE A Sequence-Indexed Library of Insertion Mutations in the Arabidopsis Genome
 JOURNAL Unpublished (2001)
 COMMENT Contact: Joseph R. Ecker
 Salk Institute Genomic Analysis Laboratory (SIGnAL)
 The Salk Institute for Biological Studies
 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
 Tel: 858 453 4100 x1752
 Fax: 858 453 6379
 Email: ecker@salk.edu

This is single pass sequence recovered from the left border of TDNA. This sequence lies within an annotated exon of At1g58150.
 Class: TDNA tagged.

FEATURES source
 1. 469
 /organism="Arabidopsis thaliana"
 /strain="Columbia 0"
 /db_xref="taxon:3702"
 /clone="SALK_011461"
 /note="PCR was performed on Arabidopsis thaliana TDNA insertion lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at <http://signal.salk.edu/tDNA-protocols.html>"

BASE COUNT 31 a 21 c 31 g 28 t
 ORIGIN

Query Match 4.28; Score 43.4; DB 17; Length 111;
 Best Local Similarity 64.4%; Pred. No. 0.018; Indels 0; Gaps 0;
 Matches 65; Conservative 0; Mismatches 36;

Qy 1 CTTACTATAGGGCACCGCTGGTCGAGGCCGGCTGGTCAATTCAAGTGGAACCTCACTG 60
 Db 105 CTCACATAGGGCACCGTGGTCGAGGCCGGCTGGTCAATTCAAGTGGAACCTCACTT 46

Qy 61 GATCCATACTGTGAGAGATAACATGACAATTCAGAGC 101
 Db 45 TATTTGAAATACCGGGCATTAACACATGGCCCTATC 5

the site of insertion. Details of the protocols used can be found at <http://signal.salk.edu/tDNA-protocols.html>"
 Class: TDNA tagged.

FEATURES source
 1. 388
 /organism="Arabidopsis thaliana"
 /strain="Columbia 0"
 /db_xref="taxon:3702"
 /clone="SALK_018023_45.85.x"
 /note="PCR was performed on Arabidopsis thaliana lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at <http://signal.salk.edu/tDNA-protocols.html>"
 3 others

BASE COUNT 93 a 120 c 81 g 135 t 40 others
 ORIGIN

Query Match 4.28; Score 43.6; DB 17; Length 469;
 Best Local Similarity 84.5%; Pred. No. 0.028; Indels 0; Gaps 0;
 Matches 49; Conservative 0; Mismatches 9;

Qy 1 CTTACTATAGGGCACCGCTGGTCGAGGCCGGCTGGTCAATTCAAGTGGAACCTCACTC 58
 Db 202 CTCACATAGGGCACGGTGGTCAGGCCGGCTGGTCAATTCAAGTGGAGCTTANC 145

RESULT 3
 LOCUS BH617520/c 111 bp DNA linear GSS 30-JAN-2002
 DEFINITION SALK_037200 Arabidopsis thaliana TDNA insertion lines Arabidopsis thaliana genomic clone SALK_037200, DNA sequence.

ACCESION BH617520
 VERSION BH617520.1 GI:18427615

ORGANISM Arabidopsis thaliana
 SOURCE thale cress.
 KEYWORDS GSS.

REFERENCE Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P., Zimmerman,J., and Ecker,J.R.

TITLE A Sequence-Indexed Library of Insertion Mutations in the Arabidopsis Genome
 JOURNAL Unpublished (2001)
 COMMENT Contact: Joseph R. Ecker
 Salk Institute Genomic Analysis Laboratory (SIGnAL)
 The Salk Institute for Biological Studies
 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
 Tel: 858 453 4100 x1752
 Fax: 858 453 6379
 Email: ecker@salk.edu

This is single pass sequence recovered from the left border of TDNA. This sequence lies within an annotated exon of At1g58150.
 Class: TDNA tagged.

FEATURES source
 1. 111
 /organism="Arabidopsis thaliana"
 /strain="Columbia 0"
 /db_xref="taxon:3702"
 /clone="SALK_037200"
 /clone.lib="Arabidopsis thaliana TDNA insertion lines"
 /note="PCR was performed on Arabidopsis thaliana TDNA insertion lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at <http://signal.salk.edu/tDNA-protocols.html>"

BASE COUNT 31 a 21 c 31 g 28 t
 ORIGIN

Query Match 4.28; Score 43.4; DB 17; Length 111;
 Best Local Similarity 64.4%; Pred. No. 0.018; Indels 0; Gaps 0;
 Matches 65; Conservative 0; Mismatches 36;

Qy 1 CTTACTATAGGGCACCGCTGGTCGAGGCCGGCTGGTCAATTCAAGTGGAACCTCACTG 60
 Db 105 CTCACATAGGGCACCGTGGTCGAGGCCGGCTGGTCAATTCAAGTGGAACCTCACTT 46

Qy 61 GATCCATACTGTGAGAGATAACATGACAATTCAGAGC 101
 Db 45 TATTTGAAATACCGGGCATTAACACATGGCCCTATC 5

RESULT 4	BH613783/c	LOCUS	BH613783 124 bp DNA linear GSS_04-JAN-2002	AUTHORS	Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P.
DEFINITION	SALK_034950 Arabidopsis thaliana TDNA insertion lines	JOURNAL	Unpublished (2001)	TITLE	A Sequence-Indexed Library of Insertion Mutations in the Arabidopsis Genome
ACCESSION	BH613783	COMMENT	Contact: Joseph R. Ecker Salk Institute Genomic Analysis Laboratory (SIGnAL) The Salk Institute for Biological Studies 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA	COMMENT	Zimmerman,J. and Ecker,J.R. Email: ecker@salk.edu
VERSION	BH613783.1 GI:18063245	FEATURES	source	Class: TDNA tagged.	This is single pass sequence recovered from the left border of TDNA.
KEYWORDS	thale cress Arabidopsis thaliana Eukaryota; Viriplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.	source		Location/Qualifiers	1. .384 /organism="Arabidopsis thaliana" /strain="Columbia 0" /db_xref="TAIR:3102" /clone_id="SALK_010743"
ORGANISM		COMMENT	Contact: Joseph R. Ecker Salk Institute Genomic Analysis Laboratory (SIGnAL) The Salk Institute for Biological Studies 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA	COMMENT	/note="PCR was performed on Arabidopsis thaliana lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at http://signal.salk.edu/tDNA_protocols.html"
REFERENCE	Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P.	BASE COUNT	87 a	ORIGIN	87 a 127 c 51 g 119 t
AUTHORS		FEATURES		Query Match	4.2%; Score 42.8; DB 17; Length 384;
JOURNAL		source		Best Local Similarity	75.7%; Pred. No. 0 046;
COMMENT				Matches	0; Mismatches 17; Indels 0; Gaps 0;
ORGANISM				DEFINITION	
REFERENCE				ACCESSION	CTTACTATAAGGCACGCCGTGTCAGGGCGGGCTGGTAGAACCTCACTG 60
AUTHORS				VERSION	106 CTCACTATAGGCACCCGTGTCAGGGCGGGCTGGCAATTGCTCTAATG 47
JOURNAL				KEYWORDS	
COMMENT				ORGANISM	Arabidopsis thaliana
ORGANISM				REFERENCE	Eukaryota; Viriplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE				AUTHORS	1 (bases 1 to 445)
AUTHORS				JOURNAL	BH634101/C
JOURNAL				COMMENT	Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P.
COMMENT				COMMENT	A Sequence-Indexed Library of Insertion Mutations in the Arabidopsis Genome
ORGANISM				UNPUBLISHED	Unpublished (2001)
REFERENCE				CONTACT	Contact: Joseph R. Ecker
DEFINITION				JOURNAL	The Salk Institute for Biological Studies
ACCESSION	BH250931/c	LOCUS	BH250931 384 bp DNA linear GSS_28-NOV-2001	COMMENT	10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
VERSION	BH250931	DEFINITION	SALK_010743 Arabidopsis thaliana TDNA insertion lines	TITLE	Unpublished (2001)
KEYWORDS		KEYWORDS		COMMENT	Contact: Joseph R. Ecker
ORGANISM		ORGANISM		JOURNAL	The Salk Institute for Biological Studies
REFERENCE		REFERENCE		COMMENT	10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
				COMMENT	Unpublished (2001)
				CONTACT	Contact: Joseph R. Ecker
				JOURNAL	The Salk Institute for Biological Studies
				COMMENT	10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
				COMMENT	Unpublished (2001)
				CONTACT	Contact: Joseph R. Ecker
				JOURNAL	The Salk Institute for Biological Studies
				COMMENT	10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
				COMMENT	Unpublished (2001)
				CONTACT	Contact: Joseph R. Ecker
				JOURNAL	The Salk Institute for Biological Studies
				COMMENT	10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
				COMMENT	Unpublished (2001)
				CONTACT	Contact: Joseph R. Ecker
				JOURNAL	The Salk Institute for Biological Studies
				COMMENT	10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
				COMMENT	Unpublished (2001)
				CONTACT	Contact: Joseph R. Ecker
				JOURNAL	The Salk Institute for Biological Studies
				COMMENT	10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
				COMMENT	Unpublished (2001)
				CONTACT	Contact: Joseph R. Ecker
				JOURNAL	The Salk Institute for Biological Studies
				COMMENT	10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
				COMMENT	Unpublished (2001)
				CONTACT	Contact: Joseph R. Ecker
				JOURNAL	The Salk Institute for Biological Studies
				COMMENT	10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
				COMMENT	Unpublished (2001)
				CONTACT	Contact: Joseph R. Ecker
				JOURNAL	The Salk Institute for Biological Studies
				COMMENT	10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
				COMMENT	Unpublished (2001)
				CONTACT	Contact: Joseph R. Ecker
				JOURNAL	The Salk Institute for Biological Studies
				COMMENT	10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
				COMMENT	Unpublished (2001)
				CONTACT	Contact: Joseph R. Ecker
				JOURNAL	The Salk Institute for Biological Studies
				COMMENT	10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
				COMMENT	Unpublished (2001)
				CONTACT	Contact: Joseph R. Ecker
				JOURNAL	The Salk Institute for Biological Studies
				COMMENT	10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
				COMMENT	Unpublished (2001)
				CONTACT	Contact: Joseph R. Ecker
				JOURNAL	The Salk Institute for Biological Studies
				COMMENT	10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
				COMMENT	Unpublished (2001)
				CONTACT	Contact: Joseph R. Ecker
				JOURNAL	The Salk Institute for Biological Studies
				COMMENT	10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
				COMMENT	Unpublished (2001)
				CONTACT	Contact: Joseph R. Ecker
				JOURNAL	The Salk Institute for Biological Studies
				COMMENT	10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
				COMMENT	Unpublished (2001)
				CONTACT	Contact: Joseph R. Ecker
				JOURNAL	The Salk Institute for Biological Studies
				COMMENT	10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
				COMMENT	Unpublished (2001)
				CONTACT	Contact: Joseph R. Ecker
				JOURNAL	The Salk Institute for Biological Studies
				COMMENT	10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
				COMMENT	Unpublished (2001)
				CONTACT	Contact: Joseph R. Ecker
				JOURNAL	The Salk Institute for Biological Studies
				COMMENT	10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
				COMMENT	Unpublished (2001)
				CONTACT	Contact: Joseph R. Ecker
				JOURNAL	The Salk Institute for Biological Studies
				COMMENT	10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
				COMMENT	Unpublished (2001)
				CONTACT	Contact: Joseph R. Ecker
				JOURNAL	The Salk Institute for Biological Studies
				COMMENT	10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
				COMMENT	Unpublished (2001)
				CONTACT	Contact: Joseph R. Ecker
				JOURNAL	The Salk Institute for Biological Studies
				COMMENT	10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
				COMMENT	Unpublished (2001)
				CONTACT	Contact: Joseph R. Ecker
				JOURNAL	The Salk Institute for Biological Studies
				COMMENT	10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
				COMMENT	Unpublished (2001)
				CONTACT	Contact: Joseph R. Ecker
				JOURNAL	The Salk Institute for Biological Studies
				COMMENT	10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
				COMMENT	Unpublished (2001)
				CONTACT	Contact: Joseph R. Ecker
				JOURNAL	The Salk Institute for Biological Studies
				COMMENT	10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
				COMMENT	Unpublished (2001)
				CONTACT	Contact: Joseph R. Ecker
				JOURNAL	The Salk Institute for Biological Studies
				COMMENT	10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
				COMMENT	Unpublished (2001)
				CONTACT	Contact: Joseph R. Ecker
				JOURNAL	The Salk Institute for Biological Studies
				COMMENT	10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
				COMMENT	Unpublished (2001)
				CONTACT	Contact: Joseph R. Ecker
				JOURNAL	The Salk Institute for Biological Studies
				COMMENT	10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
				COMMENT	Unpublished (2001)
				CONTACT	Contact: Joseph R. Ecker
				JOURNAL	The Salk Institute for Biological Studies
				COMMENT	10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
				COMMENT	Unpublished (2001)
				CONTACT	Contact: Joseph R. Ecker
				JOURNAL	The Salk Institute for Biological Studies
				COMMENT	10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
				COMMENT	Unpublished (2001)
				CONTACT	Contact: Joseph R. Ecker
				JOURNAL	The Salk Institute for Biological Studies
				COMMENT	10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
				COMMENT	Unpublished (2001)
				CONTACT	Contact: Joseph R. Ecker
				JOURNAL	The Salk Institute for Biological Studies
				COMMENT	10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
				COMMENT	Unpublished (2001)
				CONTACT	Contact: Joseph R. Ecker
				JOURNAL	The Salk Institute for Biological Studies
				COMMENT	10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
				COMMENT	Unpublished (2001)
				CONTACT	Contact: Joseph R. Ecker
				JOURNAL	The Salk Institute for Biological Studies
				COMMENT	10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
				COMMENT	Unpublished (2001)
				CONTACT	Contact: Joseph R. Ecker
				JOURNAL	The Salk Institute for Biological Studies
				COMMENT	10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
				COMMENT	Unpublished (2001)
				CONTACT	Contact: Joseph R. Ecker
				JOURNAL	The Salk Institute for Biological Studies
				COMMENT	10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
				COMMENT	Unpublished (2001)
				CONTACT	Contact: Joseph R. Ecker
				JOURNAL	The Salk Institute for Biological Studies
				COMMENT	10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
				COMMENT	Unpublished (2001)
				CONTACT	Contact: Joseph R. Ecker
				JOURNAL	The Salk Institute for Biological Studies
				COMMENT	10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
				COMMENT	Unpublished (2001)
				CONTACT	Contact: Joseph R. Ecker
				JOURNAL	The Salk Institute for Biological Studies
				COMMENT	10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
				COMMENT	Unpublished (2001)
				CONTACT	Contact: Joseph R. Ecker
				JOURNAL	The Salk Institute for Biological Studies
				COMMENT	10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
				COMMENT	Unpublished (2001)
				CONTACT	Contact: Joseph R. Ecker
				JOURNAL	The Salk Institute for Biological Studies
				COMMENT	10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
				COMMENT	Unpublished (2001)
				CONTACT	Contact: Joseph R. Ecker
				JOURNAL	The Salk Institute for Biological Studies
				COMMENT	10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
				COMMENT	Unpublished (2001)
				CONTACT	Contact: Joseph R. Ecker
				JOURNAL	The Salk Institute for Biological Studies
				COMMENT	10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
				COMMENT	Unpublished (2001)
				CONTACT	Contact: Joseph R. Ecker
				JOURNAL	The Salk Institute for Biological Studies
				COMMENT	10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
				COMMENT	Unpublished (2001)
				CONTACT	Contact: Joseph R. Ecker
				JOURNAL	The Salk Institute for Biological Studies
				COMMENT	10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
				COMMENT	Unpublished (2001)
				CONTACT	Contact: Joseph R. Ecker
				JOURNAL	The Salk Institute for Biological Studies
				COMMENT	10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
				COMMENT	Unpublished (2001)
				CONTACT	Contact: Joseph R. Ecker
				JOURNAL	The Salk Institute for Biological Studies
				COMMENT	10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
				COMMENT	Unpublished (2001)
				CONTACT	Contact: Joseph R. Ecker
				JOURNAL	The Salk Institute for Biological Studies
				COMMENT	10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
				COMMENT	Unpublished (2001)
				CONTACT	Contact: Joseph R. Ecker
				JOURNAL	The Salk Institute for Biological Studies
				COMMENT	10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
				COMMENT	Unpublished (2001)
				CONTACT	Contact: Joseph R. Ecker
				JOURNAL	The Salk Institute for Biological Studies
				COMMENT	10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
				COMMENT	Unpublished (2001)
				CONTACT	Contact: Joseph R. Ecker
				JOURNAL	The Salk Institute for Biological Studies
				COMMENT	10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
				COMMENT	Unpublished (2001)
				CONTACT	Contact: Joseph R. Ecker
				JOURNAL	The Salk Institute for Biological Studies
				COMMENT	10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
				COMMENT	Unpublished (2001)
				CONTACT	Contact: Joseph R. Ecker
				JOURNAL	The Salk Institute for Biological Studies
				COMMENT	10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
				COMMENT	Unpublished (2001)
				CONTACT	Contact: Joseph R. Ecker
				JOURNAL	The Salk Institute for Biological Studies
				COMMENT	10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
				COMMENT	Unpublished (2001)
				CONTACT	Contact: Joseph R. Ecker
				JOURNAL	The Salk Institute for Biological Studies
				COMMENT	

TITLE	A Sequence-Indexed Library of Insertion Mutations in the Arabidopsis Genome
JOURNAL	Unpublished (2001)
COMMENT	Contact: Joseph R. Ecker Salk Institute Genomic Analysis Laboratory (SIGnAL) The Salk Institute for Biological Studies 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA Tel: 858 453 4100 x1752 Fax: 858 558 6379 Email: ecker@salk.edu
This is single pass sequence recovered from the left border of TDNA.	
Class :	TDNA tagged.
FEATURES	Location/Qualifiers
source	1..129 /organism="Arabidopsis thaliana" /strain="Columbia 0" /db_xref="taxon:3702" /clone="SALK_000216" /clone_id="Arabidopsis thaliana TDNA insertion lines" /note="PCR was performed on Arabidopsis thaliana lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence a the site of insertion. Details of the protocols used be found at http://signal.salk.edu/tDNA-protocols.htm
BASE COUNT	38 c 28 g 38 t 38 a
INIGIN	

RESULT	12			
ACCSN	BR251572	BH251572	220 bp	DNA
DEFINITION	SAIK_011803	Arabidopsis thaliana	TDNA insertion lines	GSS 28-NOV-2000
SEQUENCE	SAIK_011803	Arabidopsis thaliana genomic clone	Arabidopsis	
VERSION	BH251572			
VERSION	BR251572.1			
KEYWORDS	GSS,			
SOURCE	thale cress			
ORGANISM	Arabidopsis thaliana			

REFERENCE	AUTHORS	TITLE
1	Alonso, J.M., Leisse, T.J., C., Jeske, A., Barnes, M., Zimmerman, J., and Ecker, J.R.	Barajas, P., Chen, H., Kim, C.J., Parker, H., Cheuk, R., Gadrinab, L., Prednis, L., Shinn, P.
(bases 1 to 220)		Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
		Eudicots; core eudicots; Rosidae; magnoliophyta; eudicotyledons; spermatophyta; tracheophyta; embryophyta; vascular; <i>Vitis vinifera</i> ; <i>Solanum tuberosum</i> ; <i>Arabidopsis thaliana</i> .
		A Sequence-Indexed Library of Insertion Mutations in the

JOURNAL
IMENT
nuclear genome
unpublished (2001)
Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGnAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salk.edu
this is single pass sequence recovered from the left border of
TDNA.
Class: TDNA tagged.

FEATURES	Location/Qualifiers
source	1 . 220 /organism="Arabidopsis thaliana" /strain="Columbia 0" /db_xref="taxon:3702" /clone="SALK_011803" /click="http://signal.salk.edu/tDNA-protocol?base_id=SALK_011803"
	/note="PCR was performed on Arabidopsis thaliana insertion elements. The resultant fragment for each line directly sequenced to determine the genomic sequence of insertion. Details of the protocols can be found at http://signal.salk.edu/tDNA-protocol"
BASE COUNT	44 a
ORIGIN	35 c
Query	Query Match
	Best Local Similarity
	Matches 54 ; Conservative 0; Mismatches 20; Indels 0;
QY	1 CTTACTATAGGGCATGGCTGGTCGACGCCGGCGCTGGTATGAAGGTGGGAACCTCACTA
Db	214 CTCACTATAGGCACGGCTGGTCGACGCCGGCTGGTCAAGTCAAGTCTAAACACAAT
QY	61 GATGCCATACTGC 74
Db	154 TTTTAATATAGTC 141
RESULT	RESULT 1.3
LOCUS	BH212774 /C
DEFINITION	SALK_008030 Arabidopsis thaliana TDNA insertion clone BH212774
ACCESSION	BH212774
VERSION	GI:16394148
KEYWORDS	GSS.
SOURCE	thale cress, Arabidopsis thaliana
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophytes; Magnoliophyta; eudicotyledons; core eudicots; eurosids II; Brassicales; Brassicaceae; Parker, R., Cheuk, R., C., Jeske, A., Karnes, M., Kim, C.J., Parker, R., Prednis, U., Zimmerman, J. and Ecker, J. R.
TITLE	A Sequence-indexed Library of Insertion Mutations in the Arabidopsis Genome
JOURNAL	Unpublished (2001)
COMMENT	Contact: Joseph R. Ecker Salk Institute Genomic Analysis Laboratory (SIGNAL) The Salk Institute for Biological Studies 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA tel: 858 453 4100 x1752 Fax: 858 558 6379 Email: ecker@salk.edu
	This is single pass sequence recovered from the left border TDNA.
CLASS	TDNA tagged.
FEATURES	Location/Qualifiers
source	1 . 452 /organism="Arabidopsis thaliana" /strain="Columbia 0" /db_xref="taxon:3702" /clone="SALK_008030" /click="http://signal.salk.edu/tDNA-protocol?base_id=SALK_008030"
	/note="PCR was performed on Arabidopsis thaliana insertion elements. The resultant fragment for each line directly sequenced to determine the genomic sequence of insertion. Details of the protocols can be found at http://signal.salk.edu/tDNA-protocol"
BASE COUNT	114 a
	111 c
	112 g
	112 t

